

VISIT...

LANZAROTE
Caliente.COM

THE
CAMBRIDGE
WORLD
HISTORY
OF FOOD

CAMBRIDGE
UNIVERSITY PRESS
www.cambridge.org



9 780521 402164

The
CAMBRIDGE
WORLD HISTORY
of
FOOD



Board of Editors

Kenneth F. Kiple
Kriemhild Coneè Ornelas

General Editors

George Armelagos
Department of Anthropology
Emory University
Atlanta, Georgia

Maurice Aymard
Maison des Sciences de l'Homme
Paris, France

Thomas G. Benedek
Department of Medicine
University of Pittsburgh School of
Medicine
Pittsburgh, Pennsylvania

Donald Brothwell
Institute of Archaeology
University of London
London, England

William F. Bynum
Wellcome Institute for the History of
Medicine
London, England

Doris Howes Calloway
Department of Nutritional Sciences
University of California, Berkeley
Berkeley, California

Kenneth J. Carpenter
Department of Nutritional Sciences
University of California, Berkeley
Berkeley, California

Alfred W. Crosby
Department of American Studies
University of Texas
Austin, Texas

Philip D. Curtin
Department of History
Johns Hopkins University
Baltimore, Maryland

Frederick L. Dunn
Department of Epidemiology and
Biostatistics
University of California
San Francisco, California

Stanley L. Engerman
Department of Economics and History
University of Rochester
Rochester, New York

Antoinette Fauve-Chamoux
Commission Internationale de
Démographie Historique
Paris, France

Robert Fogel
Center for Population Economics
University of Chicago
Chicago, Illinois

Daniel W. Gade
Department of Geography
University of Vermont
Burlington, Vermont

Alan H. Goodman
School of Natural Sciences
Hampshire College
Amherst, Massachusetts

Louis E. Grivetti
Department of Nutrition
University of California, Davis
Davis, California

Jerome Handler
Virginia Foundation for the Humanities
Charlottesville, Virginia

Mary Karasch
Department of History
Oakland University
Rochester, Michigan

Jack Ralph Kloppenburg, Jr.
College of Agriculture and Life Sciences
University of Wisconsin
Madison, Wisconsin

John Komlos
Seminar für Wirtschaftsgeschichte
University of Munich
Munich, Germany

Norman Kretchmer
Department of Nutritional Sciences
University of California, Berkeley
Berkeley, California

Stephen J. Kunitz
Department of Preventive Medicine
University of Rochester Medical Center
Rochester, New York

Clark Spencer Larsen
Department of Anthropology
University of North Carolina
Chapel Hill, North Carolina

Leslie Sue Lieberman
Department of Anthropology
University of Florida
Gainesville, Florida

Ellen Messer
World Hunger Program
Brown University
Providence, Rhode Island

Brian Murton
Department of Geography
University of Hawaii
Manoa, Hawaii

Marion Nestle
Department of Nutrition, Food and Hotel
Management
New York University
New York, New York

James L. Newman
Department of Geography
Syracuse University
Syracuse, New York

K. David Patterson†
Department of History
University of North Carolina
Charlotte, North Carolina

Jeffery Pilcher
Department of History
The Citadel
Charleston, South Carolina

Ted A. Rathbun
Department of Anthropology
University of South Carolina
Columbia, South Carolina

Clark Sawin
Medical Center
Veterans Administration
Boston, Massachusetts

Roger Schofield
Cambridge Group for the History of
Population and Social Structure
Cambridge, England

Frederick J. Simoons
Department of Geography
University of California, Davis
Davis, California

Noel W. Solomons
Center for Studies of Sensory Impairment,
Aging and Metabolism (CeSSIAM)
Eye and Ear Hospital
Guatemala City, Guatemala

John C. Super
Department of History
West Virginia University
Morgantown, West Virginia

Douglas H. Ubelaker
Department of Anthropology
National Museum of Natural History
Smithsonian Institution
Washington D.C.

EDITORS

Kenneth F. Kiple
Kriemhild Coneè Ornelas

EXECUTIVE EDITOR

Stephen V. Beck

ASSOCIATE EDITORS

Rachael Rockwell Graham
H. Micheal Tarver

ASSISTANT EDITORS

Jack G. Bengé
Paul Buckingham
Anne Calahan
Kristine Dahm
Julie Rae Fenstermaker
Peter Genovese
Jeffery Grim
David Harold
Carrie R. Kiple

Graham K. Kiple
Jane D. Kiple
Jonicka Peters
Shimale Robinson
Roy Smith
Jeffery Sodergren
Kerry Stewart
David Trevino
Gerald Vidro-Valentin

The
CAMBRIDGE
WORLD HISTORY
of
FOOD

EDITORS

Kenneth F. Kiple
Kriemhild Coneè Ornelas

VOLUME ONE



CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE UNIVERSITY PRESS

Cambridge, New York, Melbourne, Madrid, Cape Town, Singapore, São Paulo

Cambridge University Press

The Edinburgh Building, Cambridge CB2 2RU, United Kingdom

Published in the United States of America by Cambridge University Press, New York

www.cambridge.org

Information on this title: www.cambridge.org/9780521402163

© Cambridge University Press 2000

This book is in copyright. Subject to statutory exception and to the provision of relevant collective licensing agreements, no reproduction of any part may take place without the written permission of Cambridge University Press.

First published in print format 2000

ISBN-13 978-0-511-07413-4 eBook (Gale)

ISBN-10 0-511-07413-1 eBook (Gale)

ISBN-13 978-0-521-40216-3 hardback

ISBN-10 0-521-40216-6 hardback

Cambridge University Press has no responsibility for the persistence or accuracy of URLs for external or third-party internet websites referred to in this book, and does not guarantee that any content on such websites is, or will remain, accurate or appropriate.

The following illustrations in Part II are from the LuEsther T. Mertz Library, The New York Botanical Garden, Bronx, New York: Corn, Sorghum.

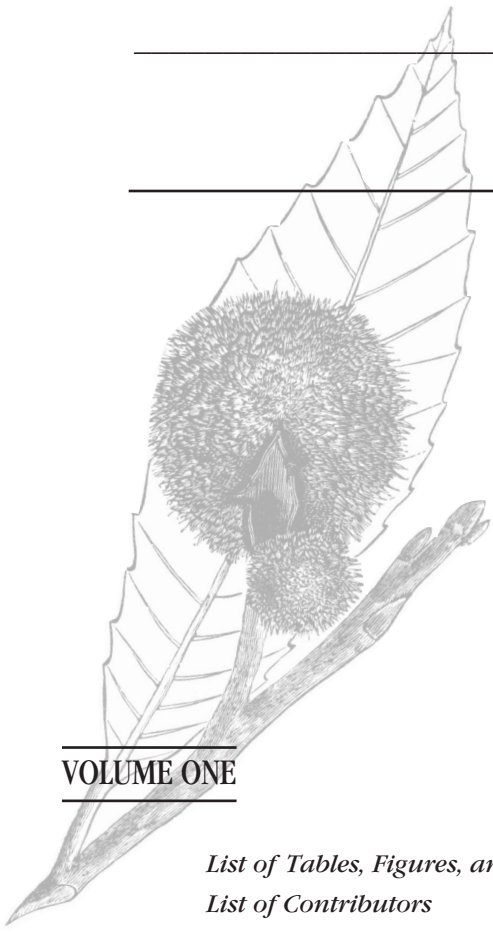
The following illustrations in Parts II and III are from the General Research Division, The New York Public Library, Astor, Lenox and Tilden Foundations: Banana plant, White potato, Prickly sago palm, Taro, Early onion, Lentil, Cabbage, Brussels sprouts, Cucumber, Watermelon, Field mushroom, Long white squash, Tomato, Chestnut, Peanut, Sesame, Soybean, Coriander, Peking duck, Geese, Goat, Cacao, Kola.

The following illustrations in Parts II and III are from the Rare Book and Manuscript Library, Columbia University: Oat, Olive, Sugar, Reindeer, Cattle, Turkey, Coffee.

In Memory of
Norman Kretchmer
Richard P. Palmieri
James J. Parsons
Daphne A. Roe
and
K. David Patterson



CONTENTS



VOLUME ONE

<i>List of Tables, Figures, and Maps</i>	<i>page</i> xix
<i>List of Contributors</i>	xxix
<i>Preface</i>	xxxvii
<i>Acknowledgments</i>	xxxix
<i>Introduction</i>	1
 Part I Determining What Our Ancestors Ate	 11
I.1. Dietary Reconstruction and Nutritional Assessment of Past Peoples: The Bioanthropological Record <i>Clark Spencer Larsen</i>	13
I.2. Paleopathological Evidence of Malnutrition <i>Donald J. Ortner and Gretchen Theobald</i>	34
I.3. Dietary Reconstruction As Seen in Coprolites <i>Kristin D. Sobolik</i>	44
I.4. Animals Used for Food in the Past: As Seen by Their Remains Excavated from Archaeological Sites <i>Elizabeth S. Wing</i>	51
I.5. Chemical Approaches to Dietary Representation <i>Ted A. Rathbun</i>	58
I.6. History, Diet, and Hunter-Gatherers <i>Mark Nathan Cohen</i>	63

Part II Staple Foods: Domesticated Plants and Animals

II.A. Grains	75
II.A.1. Amaranth <i>Mary Karasch</i>	75
II.A.2. Barley <i>Joy McCorriston</i>	81
II.A.3. Buckwheat <i>G. Mazza</i>	90
II.A.4. Maize <i>Ellen Messer</i>	97
II.A.5. Millets <i>J. M. J. de Wet</i>	112
II.A.6. Oat <i>David M. Peterson and J. Paul Murphy</i>	121
II.A.7. Rice <i>Te-Tzu Chang</i>	132
II.A.8. Rye <i>Hansjörg Küster</i>	149
II.A.9. Sorghum <i>J. M. J. de Wet</i>	152
II.A.10. Wheat <i>Joy McCorriston</i>	158
II.B. Roots, Tubers, and Other Starchy Staples	
II.B.1. Bananas and Plantains <i>Will C. McClatchey</i>	175
II.B.2. Manioc <i>Mary Karasch</i>	181
II.B.3. Potatoes (White) <i>Ellen Messer</i>	187
II.B.4. Sago <i>H. Micheal Tarver and Allan W. Austin</i>	201
II.B.5. Sweet Potatoes and Yams <i>Patricia J. O'Brien</i>	207
II.B.6. Taro <i>Nancy J. Pollock</i>	218
II.C. Important Vegetable Supplements	
II.C.1. Algae <i>Sbeldon Aaronson</i>	231
II.C.2. The <i>Allium</i> Species (Onions, Garlic, Leeks, Chives, and Shallots) <i>Julia Peterson</i>	249
II.C.3. Beans, Peas, and Lentils <i>Lawrence Kaplan</i>	271

II.C.4.	Chilli Peppers <i>Jean Andrews</i>	281
II.C.5.	Cruciferous and Green Leafy Vegetables <i>Robert C. Field</i>	288
II.C.6.	Cucumbers, Melons, and Watermelons <i>David Maynard and Donald N. Maynard</i>	298
II.C.7.	Fungi <i>Sheldon Aaronson</i>	313
II.C.8.	Squash <i>Deena S. Decker-Walters and Terrence W. Walters</i>	335
II.C.9.	Tomatoes <i>Janet Long</i>	351
II.D.	Staple Nuts	
II.D.1.	Chestnuts <i>Antoinette Fauve-Chamoux</i>	359
II.D.2.	Peanuts <i>Jobanna T. Dwyer and Ritu Sandhu</i>	364
II.E.	Animal, Marine, and Vegetable Oils	
II.E.1.	An Overview of Oils and Fats, with a Special Emphasis on Olive Oil <i>Sean Francis O'Keefe</i>	375
II.E.2.	Coconut <i>Hugh C. Harries</i>	388
II.E.3.	Palm Oil <i>K. G. Berger and S. M. Martin</i>	397
II.E.4.	Sesame <i>Dorothea Bedigian</i>	411
II.E.5.	Soybean <i>Thomas Sorosiak</i>	442
II.E.6.	Sunflower <i>Charles B. Heiser, Jr.</i>	427
II.F.	Trading in Tastes	
II.F.1.	Spices and Flavorings <i>Hansjörg Küster</i>	431
II.F.2.	Sugar <i>J. H. Galloway</i>	437
II.G.	Important Foods from Animal Sources	
II.G.1.	American Bison <i>J. Allen Barksdale</i>	450
II.G.2.	Aquatic Animals <i>Colin E. Nash</i>	456
II.G.3.	Camels <i>Elizabeth A. Stephens</i>	467

II.G.4.	Caribou and Reindeer <i>David R. Yesner</i>	480
II.G.5.	Cattle <i>Daniel W. Gade</i>	489
II.G.6.	Chickens <i>Roger Blench and Kevin C. MacDonald</i>	496
II.G.7.	Chicken Eggs <i>William J. Stadelman</i>	499
II.G.8.	Dogs <i>Stanley J. Olsen</i>	508
II.G.9.	Ducks <i>Rosemary Luff</i>	517
II.G.10.	Game <i>Stephen Beckerman</i>	524
II.G.11.	Geese <i>Kevin C. MacDonald and Roger Blench</i>	529
II.G.12.	Goats <i>Daniel W. Gade</i>	531
II.G.13.	Hogs (Pigs) <i>Daniel W. Gade</i>	536
II.G.14.	Horses <i>Daniel W. Gade</i>	542
II.G.15.	Insects <i>Darna L. Dufour and Joy B. Sander</i>	546
II.G.16.	Llamas and Alpacas <i>Daniel W. Gade</i>	555
II.G.17.	Muscovy Ducks <i>Daniel W. Gade</i>	559
II.G.18.	Pigeons <i>Richard F. Johnston</i>	561
II.G.19.	Rabbits <i>Peter R. Cheeke</i>	565
II.G.20.	Sea Turtles and Their Eggs <i>James J. Parsons</i>	567
II.G.21.	Sheep <i>Daniel W. Gade</i>	574
II.G.22.	Turkeys <i>Stanley J. Olsen</i>	578
II.G.23.	Water Buffalo <i>Robert Hoffpauir</i>	583
II.G.24.	Yak <i>Richard P. Palmieri</i>	607

Part III Dietary Liquids

III.1.	Beer and Ale <i>Phillip A. Cantrell II</i>	619
III.2.	Breast Milk and Artificial Infant Feeding <i>Antoinette Fauve-Chamoux</i>	626
III.3.	Cacao <i>Murdo J. MacLeod</i>	635
III.4.	Coffee <i>Steven C. Topik</i>	641
III.5.	Distilled Beverages <i>James Comer</i>	653
III.6.	Kava <i>Nancy J. Pollock</i>	664
III.7.	Khat <i>Clarke Brooke</i>	671
III.8.	Kola Nut <i>Edmund Abaka</i>	684
III.9.	Milk and Dairy Products <i>Keith Vernon</i>	692
III.10.	Soft Drinks <i>Colin Emmins</i>	702
III.11.	Tea <i>John H. Weisburger and James Comer</i>	712
III.12.	Water <i>Christopher Hamlin</i>	720
III.13.	Wine <i>James L. Newman</i>	730

Part IV The Nutrients – Deficiencies, Surfeits, and Food-Related Disorders**IV.A. Vitamins**

IV.A.1.	Vitamin A <i>George Wolf</i>	741
IV.A.2.	Vitamin B Complex: Thiamine, Riboflavin, Niacin, Pantothenic Acid, Pyridoxine, Cobalamin, Folic Acid <i>Daphne A. Roe</i>	750
IV.A.3.	Vitamin C <i>R. E. Hugbes</i>	754
IV.A.4.	Vitamin D <i>Glenville Jones</i>	763
IV.A.5.	Vitamin E <i>Glenville Jones</i>	769
IV.A.6.	Vitamin K and Vitamin K-Dependent Proteins <i>Myrtle Thierry-Palmer</i>	774

IV.B. Minerals

IV.B.1.	Calcium <i>Herta Spencer</i>	785
IV.B.2.	Iodine and Iodine-Deficiency Disorders <i>Basil S. Hetzel</i>	797
IV.B.3.	Iron <i>Susan Kent and Patricia Stuart-Macadam</i>	811
IV.B.4.	Magnesium <i>Theodore D. Mountokalakis</i>	824
IV.B.5.	Phosphorus <i>John J. B. Anderson</i>	834
IV.B.6.	Potassium <i>David S. Newman</i>	843
IV.B.7.	Sodium and Hypertension <i>Thomas W. Wilson and Clarence E. Grim</i>	848
IV.B.8.	Other Trace Elements <i>Forrest H. Nielsen</i>	856
IV.B.9.	Zinc <i>Ananda S. Prasad</i>	868

IV.C. Proteins, Fats, and Essential Fatty Acids

IV.C.1.	Essential Fatty Acids <i>Jacqueline L. Dupont</i>	876
IV.C.2.	Proteins <i>Kenneth J. Carpenter</i>	882
IV.C.3.	Energy and Protein Metabolism <i>Peter L. Pellett</i>	888

IV.D. Deficiency Diseases

IV.D.1.	Beriberi <i>Frederick L. Dunn</i>	914
IV.D.2.	Iron Deficiency and Anemia of Chronic Disease <i>Susan Kent</i>	919
IV.D.3.	Keshan Disease <i>Yiming Xia</i>	939
IV.D.4.	Osteoporosis <i>Robert P. Heaney</i>	947
IV.D.5.	Pellagra <i>Daphne A. Roe and Stephen V. Beck</i>	960
IV.D.6.	Pica <i>Margaret J. Weinberger</i>	967
IV.D.7.	Protein-Energy Malnutrition <i>J. D. L. Hansen</i>	977
IV.D.8.	Scurvy <i>R. E. Hughes</i>	988

IV.E. Food-Related Disorders

IV.E.1.	Anorexia Nervosa <i>Heather Munro Prescott</i>	1001
IV.E.2.	Celiac Disease <i>Donald D. Kasarda</i>	1008
IV.E.3.	Food Allergies <i>Susan L. Hefle</i>	1022
IV.E.4.	Food-Borne Infection <i>Sujatha Panikker</i>	1031
IV.E.5.	Food Sensitivities: Allergies and Intolerances <i>Judy Perkin</i>	1048
IV.E.6.	Lactose Intolerance <i>K. David Patterson</i>	1057
IV.E.7.	Obesity <i>Leslie Sue Lieberman</i>	1062

IV.F. Diet and Chronic Disease

IV.F.1.	Diabetes <i>Leslie Sue Lieberman</i>	1078
IV.F.2.	Nutrition and Cancer <i>Robert Kroes and J. H. Weisburger</i>	1086
IV.F.3.	Nutrition and Heart-Related Diseases <i>Melissa H. Olken and Joel D. Howell</i>	1097
IV.F.4.	The Cardiovascular System, Coronary Artery Disease, and Calcium: A Hypothesis <i>Stephen Seely</i>	1109

VOLUME TWO

Part V Food and Drink around the World

V.A.	The Beginnings of Agriculture: The Ancient Near East and North Africa <i>Naomi F. Miller and Wilma Wetterstrom</i>	1123
V.B.	The History and Culture of Food and Drink in Asia	
V.B.1.	The Middle East and South Asia <i>Delphine Roger</i>	1140
V.B.2.	Southeast Asia <i>Christine S. Wilson</i>	1151
V.B.3.	China <i>Françoise Sabban (translated by Elborg Forster)</i>	1165
V.B.4.	Japan <i>Naomichi Ishige</i>	1175
V.B.5.	Korea <i>Lois N. Magner</i>	1183

V.C.	The History and Culture of Food and Drink in Europe	
V.C.1.	The Mediterranean (Diets and Disease Prevention) <i>Marion Nestle</i>	1193
V.C.2.	Southern Europe <i>Kenneth Albala</i>	1203
V.C.3.	France <i>Eva Barlösius</i>	1210
V.C.4.	The British Isles <i>Colin Spencer</i>	1217
V.C.5.	Northern Europe – Germany and Surrounding Regions <i>Hansjörg Küster</i>	1226
V.C.6.	The Low Countries <i>Anneke H. van Otterloo</i>	1232
V.C.7.	Russia <i>K. David Patterson</i>	1240
V.D.	The History and Culture of Food and Drink in the Americas	
V.D.1.	Mexico and Highland Central America <i>John C. Super and Luis Alberto Vargas</i>	1248
V.D.2.	South America <i>Daniel W. Gade</i>	1254
V.D.3.	The Caribbean, Including Northern South America and Lowland Central America: Early History <i>William F. Keegan</i>	1260
V.D.4.	The Caribbean from 1492 to the Present <i>Jeffrey M. Pilcher</i>	1278
V.D.5.	Temperate and Arctic North America to 1492 <i>Elizabeth J. Reitz</i>	1288
V.D.6.	North America from 1492 to the Present <i>James Comer</i>	1304
V.D.7.	The Arctic and Subarctic Regions <i>Linda J. Reed</i>	1323
V.E.	The History and Culture of Food and Drink in Sub-Saharan Africa and Oceania	
V.E.1.	Africa South from the Sahara <i>James L. Newman</i>	1330
V.E.2.	Australia and New Zealand <i>Brian Murton</i>	1339
V.E.3.	The Pacific Islands <i>Nancy Davis Lewis</i>	1351
V.F.	Culinary History <i>Ellen Messer, Barbara Haber, Joyce Toomre, and Barbara Wheaton</i>	1367

Part VI History, Nutrition, and Health

VI.1.	Nutrition and the Decline of Mortality <i>John M. Kim</i>	1381
VI.2.	Nutrition and Mortality Decline: Another View <i>William Muraskin</i>	1389
VI.3.	Infection and Nutrition: Synergistic Interactions <i>Nevin S. Scrimshaw</i>	1397
VI.4.	Famine <i>Brian Murton</i>	1411
VI.5.	Height and Nutrition <i>Bernard Harris</i>	1427
VI.6.	The Nutrition of Women in the Developing World <i>Eileen Kennedy and Lawrence Haddad</i>	1439
VI.7.	Infant and Child Nutrition <i>Sara A. Quandt</i>	1444
VI.8.	Adolescent Nutrition and Fertility <i>Heather Munro Prescott</i>	1453
VI.9.	Nutrition and Mental Development <i>Donald T. Simeon and Sally M. Grantham-McGregor</i>	1457
VI.10.	Human Nutritional Adaptation: Biological and Cultural Aspects <i>H. H. Draper</i>	1466
VI.11.	The Psychology of Food and Food Choice <i>Paul Rozin</i>	1476
VI.12.	Food Fads <i>Jeffrey M. Pilcher</i>	1486
VI.13.	Food Prejudices and Taboos <i>Louis E. Grivetti</i>	1495
VI.14.	The Social and Cultural Uses of Food <i>Carole M. Couniban</i>	1513
VI.15.	Food as Aphrodisiacs and Anaphrodisiacs? <i>Thomas G. Benedek</i>	1523
VI.16.	Food as Medicine <i>J. Worth Estes</i>	1534
VI.17.	Vegetarianism <i>James C. Whorton</i>	1553
VI.18.	Vegetarianism: Another View <i>H. Leon Abrams, Jr.</i>	1564

Part VII Contemporary Food-Related Policy Issues

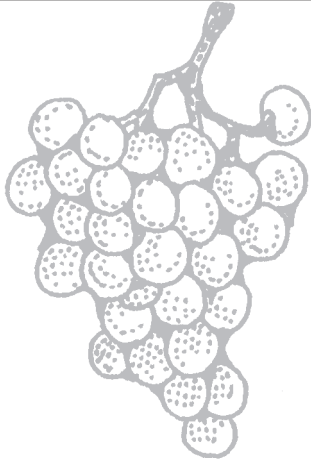
VII.1.	The State, Health, and Nutrition <i>Carol F. Helstosky</i>	1577
VII.2.	Food Entitlements <i>William H. Whitaker</i>	1585

VII.3.	Food Subsidies and Interventions for Infant and Child Nutrition <i>Penelope Nestel</i>	1593
VII.4.	Recommended Dietary Allowances and Dietary Guidance <i>Alfred E. Harper</i>	1606
VII.5.	Food Labeling <i>Eliza M. Mojdzuska</i>	1621
VII.6.	Food Lobbies and U.S. Dietary Guidance Policy <i>Marion Nestle</i>	1628
VII.7.	Food Biotechnology: Politics and Policy Implications <i>Marion Nestle</i>	1643
VII.8.	Food Safety and Biotechnology <i>Michael W. Pariza</i>	1662
VII.9.	Food Additives <i>K. T. H. Farrer</i>	1667
VII.10.	Substitute Foods and Ingredients <i>Beatrice Trum Hunter</i>	1677
VII.11.	Nonfoods as Dietary Supplements <i>R. E. Hughes</i>	1685
VII.12.	Food Toxins and Poisons from Microorganisms <i>Gordon L. Klein and Wayne R. Snodgrass</i>	1694
VII.13.	The Question of Paleolithic Nutrition and Modern Health: From the End to the Beginning <i>Kenneth F. Kiple</i>	1704

Part VIII	A Dictionary of the World's Plant Foods	1711
	Sources Consulted	1887
	Index of Latin Names	1890
	<i>Name Index</i>	1901
	<i>Subject Index</i>	1917



TABLES, FIGURES, AND MAPS



Tables

II.A.3.1.	Percent composition of buckwheat seed and its milling	page 91	II.B.6.1.	Nutritional value of the four types of taro	228
II.A.3.2.	Average mineral and vitamin contents of buckwheat whole grain	92	II.C.1.1.	Algae and blue-green bacteria eaten in contemporary Chile and Peru	232
II.A.3.3.	Chemical composition of buckwheat, barley, and corn starch granules smaller than 315 μ	92	II.C.1.2.	Algae eaten by humans now and in the past	233
II.A.3.4.	Quality of buckwheat and wheat protein	93	II.C.1.3.	The gross chemical composition of edible algae	239
II.A.3.5.	Primary grade determinants of buckwheat (Canada)	93	II.C.1.4.	Amino acid content of edible algae	240
II.A.3.6.	Absorbance of extracted color and tristimulus values of buckwheat samples stored at 25° C and 5 water activities for 19 months	94	II.C.1.5.	Vitamin content of edible algae	242
II.A.3.7.	Influence of cultivar and moisture content on dehulling characteristics and color of buckwheat seeds stored at 25° C and water activities of 0.23-0.97 for 45 days	95	II.C.1.6.	The range of fatty acids found in edible algae	244
II.A.6.1.	World oat production, area harvested, and yield by continent and country, 1965 through 1994	126	II.C.3.1.	Beans, peas, and lentils	272
II.A.7.1.	Contrast in diversification: <i>Oryza sativa</i> vs. <i>glaberrima</i>	136	II.C.6.1.	World cucumber and gherkin production, 1995	309
II.A.10.1.	Prehistoric cultures of the Near East	162	II.C.6.2.	World cantaloupe and other melon production, 1995	309
II.A.10.2.	Principal wheat types	164	II.C.6.3.	World watermelon production, 1995	310
			II.C.6.4.	Per capita consumption of cucumbers, melons, and watermelons in the United States, 1996	311
			II.C.6.5.	Nutritional composition of some cucurbits	311
			II.C.7.1.	Fungi eaten by humans around the world now and in the past	317
			II.C.7.2.	Gross chemical composition of fungi as a percentage of fungal dry weight	325

II.C.7.3.	Variations in the gross chemistry of different stages in the development of the <i>Volvariella volvacea</i> sporophore	326	II.E.1.9.	Common names for rapeseed species	383
II.C.7.4.	Vitamin content of edible fungi	326	II.E.1.10.	Fatty acid composition of butterfat	385
II.C.7.5.	Foods and beverages that require fungal processing	328	II.E.1.11.	Fat-soluble vitamin levels in fish liver oils	387
II.C.8.1.	Domesticated varieties of <i>Cucurbita argyrosperma</i> ssp. <i>argyrosperma</i>	338	II.E.3.1.	Indonesia: Oil palm area	399
II.C.8.2.	Horticultural groups of <i>Cucurbita moschata</i>	339	II.E.3.2.	Malaysia: Oil palm area	400
II.C.8.3.	Horticultural groups of <i>Cucurbita pepo</i>	341	II.E.3.3.	Latin America: Earliest oil palm plantings	400
II.C.8.4.	Horticultural groups of <i>Cucurbita maxima</i>	342	II.E.3.4.	Latin America: Oil palm area, 1992	400
II.C.8.5.	Mineral and vitamin content of young fruits, mature fruits, leaves, and growing tips and ground seed meal	349	II.E.3.5.	Comparison of village palm oil processes	401
II.D.2.1.	Characteristics of peanut varieties	365	II.E.3.6.	Palm oil exports from selected countries	406
II.D.2.2.	Comparison of various indexes of protein quality for peanuts and other protein-rich foods	368	II.E.3.7.	Specifications of special grades of crude palm oil	406
II.D.2.3.	Comparison of the amino acids in peanuts compared to high-quality proteins	368	II.E.3.8.	Palm oil imports to selected regions	407
II.D.2.4.	Comparison of nutritive value of peanuts with other common cereals and legumes	369	II.E.3.9.	Composition of palm oil	407
II.D.2.5.	Nutritional value of <i>Arachis hypogaea</i> I.	369	II.E.3.10.	Tocopherol content of typical refined palm oil	408
II.D.2.6.	Nutritive value of common peanut foods	372	II.E.6.1.	Sunflower production	429
II.E.1.1.	Average oil contents of plant sources of oil	377	II.F.1.1.	The origin of spices used historically and today	432
II.E.1.2.	Fatty acid composition ranges of natural populations of vegetable oils	379	II.G.4.1.	Weights and edible weights for caribou (<i>Rangifer tarandus</i>)	484
II.E.1.3.	Levels of minor components in olive oil	379	II.G.4.2.	Age distribution of caribou from an Alaskan archaeological site	487
II.E.1.4.	Fatty acid compositions of modified fatty acid vegetable oils	380	II.G.7.1.	Egg production of several countries of the world	502
II.E.1.5.	Tocopherol isomer distribution in dietary fats and oils	381	II.G.15.1.	Numbers of species of insects used as food by stage of life cycle and geographic region	547
II.E.1.6.	Approximate biological activity relationships of Vitamin E compounds	381	II.G.15.2.	Number of species of the insects most commonly consumed throughout the world by geographic region	552
II.E.1.7.	Nontriacylglycerol materials in crude palm oil	382	II.G.19.1.	Nutrient composition of rabbit meat	566
II.E.1.8.	Fatty acid compositions in palm olein and stearin	382	II.G.22.1.	Southwestern chronology	579
			II.G.23.1.	Population of water buffalo	586
			III.2.1.	Typical analyses of milk from various species	627
			III.2.2.	Daily quantities of milk a healthy child should ordinarily absorb during the first six months of life	627
			III.2.3.	Number of babies, aged less than 1, abandoned in Paris Foundling Hospital, 1773-7, with infant mortality for each group, according to their origin	630
			III.2.4.	General infant and child mortality in four European	

	countries during the second half of the eighteenth century	631	IV.B.4.3.	Causes of human magnesium deficiency and year of their first description	827
III.2.5.	Number of children abandoned in Paris Foundling Hospital, 1773–7, according to their age and origin	631	IV.B.4.4.	Generally accepted symptoms and signs of magnesium deficiency	829
III.7.1.	Nutritional components of khat (<i>Catha edulis</i>)	680	IV.B.4.5.	Additional symptoms and signs attributed to magnesium deficiency by some authors	829
III.8.1.	Chemical composition of the pod husk, testa, and nut of kola	685	IV.B.4.6.	Magnesium intake in the modern-day world	832
IV.A.1.1.	Association of vitamin A potency with yellow color in food	744	IV.B.5.1.	Content of phosphorus and calcium in commonly consumed foods in mg per serving	836
IV.A.1.2.	Countries categorized by degree of health; importance of vitamin A deficiency by WHO region	748	IV.C.1.1.	Unsaturated fatty acids	879
IV.A.3.1.	Ascorbic acid content of some plants	759	IV.C.2.1.	Reproduction of the final summary of the rat's requirements for amino acids, as determined by Rose and colleagues in 1948	885
IV.A.6.1.	Vitamin K analogues	775	IV.C.2.2.	The World Health Organization (1985) estimates of human requirements for protein and selected amino acids, by age	886
IV.A.6.2.	Phylloquinone content of common foods	775	IV.C.2.3.	The typical protein concentrations of a variety of foods (edible portions only) expressed as "protein calories as a percentage of total calories"	887
IV.B.1.1.	Calcium balances of males and females during a low calcium intake	788	IV.C.3.1.	Energy exchange in humans: An example from Atwater and Benedict	891
IV.B.1.2.	Studies of the calcium requirement	789	IV.C.3.2.	Food availability data for industrialized and developing countries. Data for 1994	893
IV.B.1.3.	Effect of aluminum-containing antacids on the calcium and phosphorus balance	790	IV.C.3.3.	The principal dietary carbohydrates	893
IV.B.1.4.	Effect of a high-protein diet on calcium metabolism	793	IV.C.3.4.	Equations for predicting basal metabolic rate from body weight and age	895
IV.B.1.5.	Patients with chronic alcoholism and osteoporosis	793	IV.C.3.5.	Physical activity levels suggested to estimate total daily energy expenditure from the mean basal metabolic rate of children, adolescents, and adults	897
IV.B.1.6.	Effect of corticosteroids on the calcium balance	793	IV.C.3.6.	Early protein and energy intakes from Europe and America with requirement estimates	898
IV.B.2.1.	The spectrum of iodine-deficiency disorders	803	IV.C.3.7.	International food energy requirements (1950–96)	898
IV.B.2.2.	Estimated prevalence of iodine-deficiency disorders in developing countries, by region and numbers of persons at risk	806	IV.C.3.8.	Distribution of food energy, fat, and protein in the various world regions	899
IV.B.3.1.	Normal hematological values for the more common iron indexes	814	IV.C.3.9.	Summary of nonessential amino acid biosynthesis in mammals	903
IV.B.3.2.	Comparison of laboratory values of anemia of dietary iron deficiency and anemia of chronic disease	815	IV.C.3.10.	Other functions of some amino acids	903
IV.B.3.3.	Types of disorders associated with iron overload	818			
IV.B.3.4.	Morphological classification of anemia	820			
IV.B.4.1.	Some important dates with reference to magnesium	826			
IV.B.4.2.	Year of first application of different procedures for measuring magnesium in biological materials	826			

IV.C.3.11.	Fate of the nitrogen and the carbon atoms in the degradation of the amino acids for energy	904	IV.D.3.7.	Indexes for oxidant defense capability in blood of children from Dechang and Mianning Counties in 1987	945
IV.C.3.12.	Recommended scoring patterns, 1950–91	905	IV.D.3.8.	Comparison of selenium content in cereals and human hair between the 1970s and 1980s	946
IV.C.3.13.	International protein recommendation (1936–96)	906	IV.D.4.1.	Threshold calcium intakes during growth	957
IV.C.3.14.	Summary of some recent committee recommendations for practical protein allowances in various age groups	906	IV.D.7.1.	The Wellcome classification of PEM	978
IV.C.3.15.	FAO/WHO/UNU (1985) safe levels of protein intake	907	IV.E.3.1.	Symptoms of food allergy	1023
IV.C.3.16.	Factorial approach for human protein requirements: Adapted from FAO/WHO (1973)	907	IV.E.3.2.	Common allergenic foods	1024
IV.C.3.17.	Amino acid composition of major food groups from the Massachusetts Nutrient Data Bank	908	IV.E.4.1.	Organisms causing food-borne disease	1032
IV.C.3.18.	Mean values per capita for the availability of specific indispensable amino acids in developed and developing regions. Data for 1994	908	IV.E.5.1.	The Type I allergic reaction	1049
IV.C.3.19.	A proposed classification using BMI (WT/HT ²)	909	IV.E.5.2.	Major food allergens	1050
IV.D.3.1.	Keshan disease incidence and prognosis of selenium-supplemented and control children (1–9 years old) in Mianning County during 1974–7	942	IV.E.6.1.	Distribution of lactose phenotypes	1060
IV.D.3.2.	Keshan disease incidence in selenium-supplemented and control children (1–12 years old) in five counties of Sichuan Province during 1976–80	942	IV.E.7.1.	Prevalence of overweight (1980s–90s), based on Body Mass Index or weight for height references	1063
IV.D.3.3.	Selenium levels in human blood and hair from residents in Keshan disease-affected and nonaffected areas in 1972–3	943	IV.E.7.2.	Prevalence of obesity	1064
IV.D.3.4.	Blood glutathione peroxidase (GPX) activities of children from Keshan disease-affected and nonaffected areas in 1975	943	IV.E.7.3.	Age-adjusted and age-specific prevalence of overweight (1960–91)	1064
IV.D.3.5.	Selenium contents of blood, hair, and grains in Keshan disease-affected and nonaffected areas	943	IV.F.1.1.	A historical perspective on dietary recommendations for people with diabetes	1080
IV.D.3.6.	Selenium contents and glutathione peroxidase (GPX) activities in tissues from patients with subacute Keshan disease and controls in affected or nonaffected areas	944	IV.F.2.1.	Chronic disease prevention and health promotion	1095
			IV.F.4.1.	Correlation coefficients between age-compensated male mortality rates from ischaemic heart disease and the consumption of various foods in eight member countries of the Organization of Economic Cooperation and Development	1116
			IV.F.4.2.	Sample data on which Table IV.F.4.1 is based	1116
			V.A.1.	Chronology of the Near East and Egypt	1124
			V.A.2.	Pharaonic Egypt	1132
			V.C.1.1.	Sources of information about diets in ancient Egypt	1194
			V.C.1.2.	Dietary intake in Crete in 1948 as estimated by three methods	1195
			V.C.1.3.	Percentage of total energy contributed by major food groups in the diet of Crete as compared to their availability in the food	

	supplies of Greece and the United States in 1948-9	1195	VI.11.1.	Psychological categorization of acceptance and rejection	1478
V.C.1.4.	Ancel and Margaret Keys' 1959 dietary advice for the prevention of coronary heart disease compared to the 1995 U.S. dietary guidelines	1197	VI.13.1.	Selected forbidden foods: Leviticus (Hebrew source with English translations)	1499
V.C.1.5.	Suggestions for further historical and applied research on the health impact of Mediterranean diets	1201	VI.13.2.	BaTlokwa ba Moshaweng: Foods restricted by gender and age	1505
V.C.7.1.	Indexes of food consumption by collective farm workers	1245	VI.15.1.	Ancient sexual stimulants and depressants	1523
V.C.7.2.	Consumption of major foods, 1913-76	1245	VI.15.2.	Most commonly cited aphrodisiacs, 1546-1710	1527
V.C.7.3.	Food as a percentage of family expenditure, 1940-90	1245	VI.16.1.	Past and present medicinal uses of flavorings and spices	1540
V.D.3.1.	Comparison of house-garden cultigens in native Amazonian and prehistoric West Indian gardens	1264	VI.16.2.	Past and present medicinal uses of fruits and nuts	1540
V.D.3.2.	Fishes identified in Lucayan sites	1265	VI.16.3.	Past and present medicinal uses of vegetables	1541
V.D.3.3.	Return rates and resource rankings of Lucayan foods	1267	VI.16.4.	Past and present medicinal uses of beverages	1541
V.D.3.4.	Garifuna ceremonial foods and the probable time of their introduction	1273	VI.16.5.	Past and present medicinal uses of grains	1541
V.D.5.1.	General chronological sequence	1288	VI.16.6.	Past and present medicinal uses of gums and roots	1541
V.D.5.2.	List of scientific and common names for plants	1289	VI.16.7.	Past and present medicinal uses of miscellaneous foodstuffs	1541
V.D.5.3.	List of scientific and common names for animals	1290	VII.6.1.	Selected landmarks in the history of U.S. lobbying	1631
VI.3.1.	108 acute infections among 32 children ages 2 to 9 years observed in a "model" convalescent home in Guatemala City for 90 days	1398	VII.6.2.	Selected examples of food lobbying groups	1633
VI.3.2.	Antimicrobial systems in the neutrophil	1399	VII.6.3.	A partial list of food and agriculture Political Action Committees (PACs) contributing to the 1989-90 election campaign of Senator Tom Harkin (D-IA), a member of the Appropriations and Agriculture, Nutrition and Forestry Committees	1634
VI.3.3.	Intake of calories in acute state, and 2 weeks and 8 weeks after recovery	1404	VII.6.4.	Evolution of federal recommendations to reduce dietary fat through changes in meat consumption	1637
VI.5.1.	Changes in the heights of European army recruits circa 1900-1975	1429	VII.7.1.	Theoretical and current applications of food biotechnology	1644
VI.5.2.	Median menarcheal age of girls in various European countries, 1950s-60s and 1970s-80s	1430	VII.7.2.	Key events in the history of the commercialization of food products of biotechnology in the United States	1647
VI.5.3.	Average heights of selected groups of Indo-Mediterranean children at different periods	1434	VII.7.3.	Safety issues raised by food biotechnology	1648
VI.7.1.	Percentage of first-born infants ever breast-fed between 1951 and 1970 in the United States, by ethnic group and education	1449	VII.7.4.	The principal arguments for and against the patenting of transgenic animals	1651

VII.7.5.	Public perceptions of food biotechnology	1653	I.1.13.	Cribra orbitalia in historic Florida Indian	27
VII.7.6.	Analytical framework for predicting public acceptance of a food product of biotechnology	1654	I.2.1.	External view of the maxilla of a child about 6 years of age at the time of death	37
VII.8.1.	Ranking food safety risks	1663	I.2.2.	Right sphenoid and adjacent bone surfaces of case seen in Figure I.2.1	38
VII.8.2.	Summary of reported food-borne disease outbreaks in the United States, 1983–7	1663	I.2.3.	Orbital roof of case seen in Figure I.2.1	38
VII.8.3.	Some natural pesticidal carcinogens in food	1665	I.2.4.	Inner table of the frontal bone of case seen in Figure I.2.1	38
VII.9.1.	Food additives and their functions	1671	I.2.5.	Right lateral view of the ninth through the twelfth thoracic vertebrae from the skeleton of a male about 45 years of age at the time of death	40
VII.11.1.	Publications during 1930–90 relating to the nutritional significance of bioflavonoids and carnitine	1687	I.2.6a.	Photomicrograph of a bone section from the femur of the burial seen in Figure I.2.5	41
Figures			I.2.6b.	Photomicrograph of a microradiograph of the bone section seen in Figure I.2.6a	41
I.1.1.	Temporal changes in mean values of $\delta^{13}\text{C}$ of prehistoric eastern North American Indians	17	II.A.3.1.	Flow diagram of two buckwheat mills: (A) roller mill; (B) stone-roller mill	96
I.1.2.	Scanning electron micrographs of prehistoric hunter-gatherer molar and historic agriculturalist molar from the southeastern U.S. Atlantic coast	18	II.A.6.1.	Flow diagram of typical oat-milling sequence	128
I.1.3.	Views of mandibular dentitions showing agriculturalist and hunter-gatherer wear planes	19	II.A.10.1.	Related wheats and goat-faced grasses	166
I.1.4.	Lingual wear on anterior teeth of prehistoric Brazilian Indian	19	II.A.10.2.	Photograph of the Nahal Hemar sickle	167
I.1.5.	Dental carious lesion in maxillary molar from historic Florida Indian	20	II.B.6.1.	Different types of taros	220
I.1.6.	Growth curves from Dickson Mounds, Illinois, Indian population	21	II.C.4.1.	Cross-section of a pepper	286
I.1.7.	Micrograph showing hyper-mineralized rings within an osteon from prehistoric Nubian	22	II.C.6.1.	Netted cantaloupe fruit	299
I.1.8.	Radiograph and section of prehistoric California Indian femur with Harris lines	23	II.C.6.2.	Casaba melon	300
I.1.9.	Juvenile anterior dentition showing hypoplasias on incompletely erupted incisors	24	II.C.6.3.	Juan Canary melon	301
I.1.10.	Micrograph of canine tooth showing Wilson band from Native American Libben site	25	II.C.6.4.	Santa Claus melon	301
I.1.11.	Femora and tibiae of nineteenth-century black American showing limb bone deformation due to rickets	26	II.C.6.5.	Pistillate and staminate cucumber flowers	302
I.1.12.	Porotic hyperostosis on prehistoric Peruvian Indian posterior cranium	27	II.C.6.6.	Cucumbers	302
			II.C.6.7.	Gynoecious, parthenocarpic greenhouse cucumbers	303
			II.C.6.8.	Variation in watermelon fruit size, shape, and color and flesh color	303
			II.C.6.9.	Seedless watermelon with seeded watermelon	304
			II.C.6.10.	'Jubilee' watermelon developed by J. M. Crall, University of Florida, in 1963	308
			II.C.6.11.	Watermelon seedlings grafted by machine onto Fusarium-resistant rootstocks in Japan	309

II.C.6.12.	Melons for sale as special gifts in Kyoto, Japan	310	II.G.4.1.	Seasonal variation in the fat content of caribou	483
II.C.6.13.	Low, supported row covers for watermelon production in Daiei, Japan	310	II.G.4.2.	A spring drying rack for caribou meat used by the Nunamiut Eskimo of the Brooks Range, northern Alaska	485
II.C.6.14.	NMR watermelon quality determination in Japan	312	II.G.7.1.	Structure of the chicken egg	503
II.C.6.15.	Watermelon for sale in Japan at U.S. \$50	312	II.G.8.1.	Intentional burials of domestic dogs and humans from the Neolithic in China. Xiawanggang c. 4,000 B.P.	509
II.C.8.1.	<i>Cucurbita moschata</i>	338	II.G.8.2.	Indian dogs, 2,000 years old, early Basketmaker. Natural mummies from White Dog Cave, Marsh Pass, Arizona	510
II.C.8.2.	Seeds of <i>Cucurbita pepo</i> , <i>C. moschata</i> , <i>C. argyrosperma</i> , and 'Silverseed Gourd'	339	II.G.8.3.	Typical long-legged Basketmaker domestic dogs, <i>Canis familiaris</i> , from the vicinity of Marsh Pass, Arizona, 2,000 B.P.	511
II.C.8.3.	'Butternut', a "bell squash" cultivar of <i>Cucurbita moschata</i>	340	II.G.9.1.	The mallard duck	521
II.C.8.4.	An unusual "acorn squash" of <i>Cucurbita pepo</i>	340	II.G.22.1.	Early pueblo domestic turkeys	580
II.C.8.5.	'Delicata' (<i>Cucurbita pepo</i>)	341	II.G.22.2.	The dog and the turkey: The only two domestic animals of the southwestern pueblos upon the arrival of the Europeans	582
II.C.8.6.	Various small-fruited cultivars of <i>Cucurbita pepo</i>	341	II.G.23.1.	The domesticated water buffalo	584
II.C.8.7.	'Turk's Turban' (<i>Cucurbita maxima</i>)	342	II.G.23.2.	Wild buffalo in Assam, with typical riparian and tall-grass habitat depicted	584
II.C.8.8.	'Buttercup', a "turban squash" of <i>Cucurbita maxima</i>	343	II.G.23.3.	Depictions of water buffalo on seal-amulets from Mohenjo-daro	594
II.C.8.9.	A "hubbard squash" of <i>Cucurbita maxima</i>	343	II.G.23.7.	Depictions of water buffaloes on cylinder seals from Mesopotamia	597
II.C.8.10.	Mature fruit of <i>Cucurbita argyrosperma</i> ssp. <i>sororia</i>	344	II.G.24.1.	Domesticated yak	609
II.C.8.11.	'Seminole Pumpkin' (<i>Cucurbita moschata</i>)	345	IV.A.1.1.	Chemical structure of all- <i>trans</i> -retinol and all- <i>trans</i> -beta-carotene	741
II.C.8.12.	Wild <i>Cucurbita pepo</i> ssp. <i>ovifera</i> var. <i>ozarkana</i> from a riparian site in the Mississippi Valley	346	IV.A.1.2.	A postulated mechanism for the pathogenesis of keratomalacia in vitamin A deficiency	747
II.C.8.13.	Wild spp. <i>ovifera</i> var. <i>texana</i> , 'Mandan' and wild ssp. <i>fraterna</i>	346	IV.A.6.1.	The vitamin K-dependent carboxylase reaction	777
II.E.1.1.	The structure of common sterols	376	IV.A.6.2.	An outline of the clotting sequence	778
II.E.1.2.	Production estimates for important fats and oils	376	IV.A.6.3.	The vitamin K-dependent anticoagulant system	779
II.E.1.3.	Operations in soybean oil extraction and refining	377	IV.B.2.1.	"The Reun cretin," from the Reun Model Book, produced by the Cistercian Abbey at Reun, Austria, thirteenth century	798
II.E.1.4.	Tocopherol composition of soybean oil	381	IV.B.2.2.	Madonna and child by Francesco di Gentili, fifteenth century	799
II.E.1.5.	Effects of overfishing in Pacific sardine fishery, 1918-60	386	IV.B.2.3.	A dwarfed cretin from Xingjiang China, who is also deaf-mute	802
II.E.3.1.	World production of palm oil, 1910-90	406			
II.F.2.1.	Centrifugal sugar; world production	440			
II.F.2.2.	A Caribbean sugar factory	441			
II.F.2.3.	Oxen drawing a cart of cane, with mill in background	443			
II.G.3.1.	<i>Camelus dromedarius</i>	468			
II.G.3.2.	<i>Camelus bactrianus</i>	469			

IV.B.2.4.	The results of a controlled trial of iodized oil injection in the Jimi River district of the highlands of Papua New Guinea	803	IV.D.4.3.	Relationship of calcium intake, absorption efficiency, and net absorption	953
IV.B.2.5.	Nodular goiter in a New Guinean before and three months after injection of iodized oil	807	IV.D.4.4.	Threshold behavior of calcium intake	955
IV.B.5.1.	Approximate percentage contributions of the major food groups to the consumption of phosphorus	835	IV.D.4.5.	Relationship of calcium intake to calcium balance in adolescents	956
IV.B.5.2a.	Median phosphorus and calcium consumption of females in the United States	837	IV.D.7.1.	Marasmus	978
IV.B.5.2.b.	The dietary calcium to phosphorus ratio of females across the life cycle	837	IV.D.7.2.	Kwashiorkor	980
IV.B.5.3.	Schematic diagram of phosphorus balance of an adult male	840	IV.E.2.1.	Evolutionary factors combine to produce celiac disease	1009
IV.B.5.4.	Mechanism through which a low dietary calcium:phosphorus ratio contributes to the development of a persistently elevated parathyroid hormone (PTH) concentration in the blood	841	IV.F.4.1.	Male coronary mortality in the 65–74 age group in OECD countries and the consumption of milk proteins (excluding cheese)	1117
IV.B.5.5.	Comparison of parathyroid hormone (PTH) responses of normal and high dietary phosphorus and effects of PTH on bone mass	842	V.C.7.1.	Pure alcohol consumption per person over 15 years old, 1955–79	1246
IV.B.6.1.	The Na ⁺ -K ⁺ -ATPase transmembrane pump pumping Na ⁺ ions out of the cell and K ⁺ ions into the cell	845	V.D.3.1.	Isotopic reconstruction of Lucayan consumption	1261
IV.C.1.1.	Desaturation, elongation, and chain shortening of families of unsaturated fatty acids	879	VI.3.1.	Cutaneous delayed hypersensitivity to 5 t.u. tuberculin related to serum transferrin concentration in patients with pulmonary tuberculosis	1399
IV.C.1.2.	Conversion of arachidonic acid into eicosanoids	880	VI.3.2.	Serum C3 levels correlated with infection-morbidity indicated by the number of days of fever	1400
IV.C.3.1.	The Atwater bomb calorimeter	892	VII.6.1.	Meat and dairy groups approved of the 1958 <i>Basic Four</i>	1630
IV.C.3.2.	An overview of the combustion of fuels for energy	893	VII.7.1.	Food and Drug Administration policy guidelines for regulation of foods developed through biotechnology	1948
IV.C.3.3.	Metabolism of dietary protein	902	VII.11.1.	Eighteenth-century description of scurvy which includes fatigue and lassitude	1691
IV.C.3.4.	Metabolic pathways for the amino acids	905	VII.11.2.	Another eighteenth-century description of scurvy suggestive of carnitine deficiency	1691
IV.D.3.1.	The incidence, mortality, and case-fatality of Keshan disease in China	940		Maps	
IV.D.4.1.	Causal connections of some of the major factors influencing bone strength	951	II.A.1.1.	Mexico: Localities and regions where grain amaranth cultivation is indicated	77
IV.D.4.2.	Schematic illustration of the relationship between body depletion of a nutrient and health status	952	II.A.1.2.	South America: Localities and regions where grain amaranth cultivation is indicated	79
			II.A.7.1.	Extent of wild relatives and spread of ecogeographic races of <i>O. sativa</i> in Asia and Oceania	137

II.A.10.1.	The Ancient Near East showing sites mentioned in the text	159	II.G.23.2.	Buffalo in Pleistocene and Early Holocene (Paleolithic) of southern and eastern Asia	590
II.A.10.2.	The Near East with modern "hilly flanks" and Mediterranean woodlands	160	II.G.23.3.	Recent distribution of wild buffaloes	592
II.A.10.3.	Geographic distribution of wild einkorn wheat, <i>Triticum boeoticum</i>	169	II.G.23.4.	Buffalo in Neolithic and Metal Age sites	595
II.A.10.4.	Geographic distribution of wild emmer wheat, <i>Triticum dicoccoides</i>	169	II.G.23.5.	Tribal groups practicing buffalo sacrifice	599
II.A.10.5.	Geographic distribution of goat-faced grass, <i>Aegilops tauchii</i>	170	IV.B.2.1.	The distribution of iodine-deficiency disorders in developing countries	806
II.G.3.1.	The approximate modern distribution of camels	470	IV.B.4.1.	Magnesia and its colonies in Asia Minor. The migration of <i>Magnetes</i> during the twelfth and eleventh centuries B.C.	825
II.G.3.2.	Archaeological sites mentioned in the text	477	V.A.1.	The world of Pharaonic Egypt	1131
II.G.23.1.	World distribution of water buffalo	587	V.D.5.1.	The Eastern Woodlands	1292



CONTRIBUTORS



Sheldon Aaronson
Department of Biology
Queens College - CUNY
Flushing, New York

Edmund Abaka
Department of History
University of Miami
Miami, Florida

H. Leon Abrams, Jr.
Consulting Anthropologist
Bloomfield, New Jersey

Kenneth Albala
Department of History
University of the Pacific
Stockton, California

John J. B. Anderson
Department of Nutrition
University of North Carolina
Chapel Hill, North Carolina

Jean Andrews
Department of Botany
University of Texas
Austin, Texas

Allan W. Austin
Department of History
University of Cincinnati
Cincinnati, Ohio

J. Allen Barksdale
American Culture Studies
Bowling Green State University
Bowling Green, Ohio

Eva Barlösius
Institut für Agrarpolitik, Marktforschung und
Wirtschaftssoziologie der Universität Bonn
Bonn, Germany

Stephen V. Beck
Department of History
Bowling Green State University
Bowling Green, Ohio

Stephen Beckerman
Department of Anthropology
Pennsylvania State University
University Park, Pennsylvania

Dorothea Bedigian
Antioch College
Yellow Springs, Ohio

Thomas G. Benedek
Department of Medicine
University of Pittsburgh School of Medicine
Pittsburgh, Pennsylvania

K. G. Berger
Technical Consultant - Oils and Fats
Chiswick
London, England

Roger Blench
Overseas Development Institute
London, England

Clarke Brooke
Department of Geography
Portland State University
Portland, Oregon

Phillip A. Cantrell, III
Department of History
West Virginia University
Morgantown, West Virginia

Kenneth J. Carpenter
Department of Nutritional Sciences
University of California, Berkeley
Berkeley, California

Te-Tzu Chang
International Rice Research Institute
Tamshui
Taipei, Taiwan

Peter R. Cheeke
Department of Animal Sciences
Oregon State University
Corvallis, Oregon

Mark N. Cohen
Department of Anthropology
State University of New York
Plattsburgh, New York

James Comer
Department of History
Bowling Green State University
Bowling Green, Ohio

Carole M. Counihan
Department of Sociology and Anthropology
Millersville University of Pennsylvania
Millersville, Pennsylvania

Deena S. Decker-Walters
The Cucurbit Network
P.O. Box 560483
Miami, Florida

J. M. J. de Wet
University of Illinois Champaign-Urbana
Urbana, Illinois

Harold H. Draper
Department of Nutritional Sciences
University of Guelph
Guelph, Ontario
Canada

Darna L. Dufour
Department of Anthropology
University of Colorado at Boulder
Boulder, Colorado

Frederick L. Dunn
Department of Epidemiology and Biostatistics
University of California School of Medicine
San Francisco, California

Jacqueline L. Dupont
Department of Nutrition
Florida State University
Tallahassee, Florida

Johanna Dwyer
Frances Stern Nutrition Center
New England Medical Center
Boston, Massachusetts

Colin Emmins
Freelance writer and researcher
West Ealing
London, England

J. Worth Estes
Department of Pharmacology and Experimental
Therapeutics
Boston University School of Medicine
Boston, Massachusetts

K. T. H. Farrer
Consultant in Food Science and Technology
Chandler's Ford
Hants, England

Antoinette Fauve-Chamoux
Commission Internationale de Démographie Historique
Paris, France

Robert C. Field
Department of History
Bowling Green State University
Bowling Green, Ohio

Daniel W. Gade
Department of Geography
University of Vermont
Burlington, Vermont

J. H. Galloway
Department of Geography
University of Toronto
Toronto, Canada

Sally M. Grantham-McGregor
Institute of Child Health
University College London
London, England

Clarence E. Grim
Division of Cardiology
Medical College of Wisconsin
Milwaukee, Wisconsin

Louis E. Grivetti
Department of Nutrition
University of California, Davis
Davis, California

Barbara Haber
Curator of Books, Schlesinger Library
Radcliffe College
Cambridge, Massachusetts

Lawrence Haddad
Food Consumption and Nutrition Division
International Food Policy Research Institute
Washington, D.C.

Christopher Hamlin
Department of History
University of Notre Dame
South Bend, Indiana

John Derek Lindsell Hansen
Department of Paediatrics and Child Health
University of Witwatersrand
Johannesburg, Republic of South Africa

Alfred E. Harper
Department of Nutritional Sciences
Department of Biochemistry
University of Wisconsin-Madison
Madison, Wisconsin

Hugh C. Harries
Centro de Investigación Científica de Yucatán AC
Cordemex, Merida
Yucatán, Mexico

Bernard Harris
Department of Sociology and Social Policy
University of Southampton
Southampton, England

Robert P. Heaney
John A. Creighton University Professor
Creighton University
Omaha, Nebraska

Susan L. Hefle
Department of Food Science and Technology
University of Nebraska
Lincoln, Nebraska

Charles B. Heiser, Jr.
Department of Biology
Indiana University
Bloomington, Indiana

Carol F. Helstosky
Department of History
University of Denver
Denver, Colorado

Basil S. Hetzel
International Council for the Control of Iodine Deficiency
Disorders
Adelaide Medical Centre for Women and Children
North Adelaide, Australia

Robert Hoffpauir
Department of Geography
California State University
Northridge, California

Joel D. Howell
Clinical Scholars Program
University of Michigan
Ann Arbor, Michigan

R. Elwyn Hughes
School of Pure and Applied Biology
University of Wales at Cardiff
Cardiff, Wales

Beatrice Trum Hunter
Food Editor, *Consumer's Research*
Hillsboro, New Hampshire

Naomichi Ishige
National Museum of Ethnology
Osaka, Japan

Richard F. Johnston
Department of Biological Sciences
University of Kansas
Lawrence, Kansas

Glenville Jones
Department of Biochemistry
Queen's University
Kingston, Canada

Lawrence Kaplan
Department of Biology
University of Massachusetts
Boston, Massachusetts

Mary Karasch
Department of History
Oakland University
Rochester, Michigan

Donald D. Kasarda
U.S. Department of Agriculture
Western Regional Research Center
Albany, California

William F. Keegan
Department of Anthropology
Florida Museum of Natural History
University of Florida
Gainesville, Florida

Eileen Kennedy
International Food Policy Research Institute
Washington, D.C.

Susan Kent
Anthropology Program
Old Dominion University
Norfolk, Virginia

John M. Kim
Center for Population Studies
Graduate School of Business
University of Chicago
Chicago, Illinois

Kenneth F. Kiple
Department of History
Bowling Green State University
Bowling Green, Ohio

Gordon L. Klein
Pediatric Gastroenterology Division, Child Health Center
University of Texas Medical Branch
Galveston, Texas

Robert Kroes
Research Institute for Toxicology, Utrecht University
Utrecht, The Netherlands

Hansjörg Küster
Institut für Geobotanik
Universität Hannover
Hannover, Germany

Clark Spencer Larsen
Research Laboratory of Archaeology
Department of Anthropology
University of North Carolina
Chapel Hill, North Carolina

Nancy Davis Lewis
Department of Geography
University of Hawaii
Honolulu, Hawaii

Leslie Sue Lieberman
Department of Anthropology
University of Florida
Gainesville, Florida

Janet Long
Instituto de Investigaciones Históricas
Ciudad de la Investigación en Humanidades
Ciudad Universitaria
Mexico City, Mexico

Rosemary Luff
Department of Archaeology
University of Cambridge
Cambridge, England

Kevin C. MacDonald
Institute of Archaeology
University College London
London, England

Murdo J. MacLeod
Department of History
University of Florida
Gainesville, Florida

Lois N. Magner
Department of History
Purdue University
West Lafayette, Indiana

Susan M. Martin
Harpenden
Hertfordshire, England

David Maynard
Department of Sociology and Anthropology
Baruch College, City University of New York
New York, New York

Donald N. Maynard
Institute of Food and Agricultural Sciences
University of Florida
Bradenton, Florida

G. Mazza
Agriculture and Agri-Food Canada
Pacific Agri-Food Research Centre
Summerland, British Columbia
Canada

Will C. McClatchey
Department of Botany
University of Hawaii at Manoa
Honolulu, Hawaii

Joy McCorriston
Department of Anthropology
Ohio State University
Columbus, Ohio

Ellen Messer
World Hunger Program
Brown University
Providence, Rhode Island

Naomi F. Miller
The University Museum
University of Pennsylvania
Philadelphia, Pennsylvania

Eliza Mojduszka
Department of Resource Economics
University of Massachusetts
Amherst, Massachusetts

T. D. Mountokalakis
University of Athens Medical School
Athens, Greece

William Muraskin
Department of Urban Studies
Queens College, City University of New York
Flushing, New York

J. Paul Murphy
Department of Crop Science
North Carolina State University
Raleigh, North Carolina

Brian Murton
Department of Geography
University of Hawaii
Honolulu, Hawaii

Colin E. Nash
National Marine Fisheries Service (NOAA)
Seattle, Washington

Penelope Nestel
Demographic and Health Surveys
IRD/Macro Systems International, Inc.
Columbia, Maryland

Marion Nestle
Department of Nutrition, Food and Hotel Management
New York University
New York, New York

David S. Newman
Department of Chemistry
Bowling Green State University
Bowling Green, Ohio

James L. Newman
Department of Geography
Syracuse University
Syracuse, New York

Forrest H. Nielsen
Northern Plains Area Grand Forks Human Nutrition
Center
U.S. Department of Agriculture
Grand Forks, North Dakota

Patricia O'Brien
Department of Sociology, Anthropology, and
Social Work
Kansas State University
Manhattan, Kansas

Sean F. O'Keefe
Department of Food Sciences and Human Nutrition
University of Florida
Gainesville, Florida

Melissa Hendrix Olken
Department of Internal Medicine
St. Joseph Mercy Hospital
Ann Arbor, Michigan

Stanley J. Olsen
Department of Anthropology
University of Arizona
Tucson, Arizona

Donald J. Ortner
Department of Anthropology
National Museum of Natural History
Smithsonian Institution
Washington, D.C.

Anneke H. van Otterloo
Faculty of Political and Socio-Cultural Sciences
University of Amsterdam
Amsterdam, The Netherlands

Richard P. Palmieri†
Department of Geography
Mary Washington College
Fredericksburg, Virginia

Sujatha Panikker
Department of Medical Microbiology
University of Manchester Medical School
Manchester, England

Michael W. Pariza
Food Research Institute
Department of Food Microbiology and Toxicology
University of Wisconsin-Madison
Madison, Wisconsin

James J. Parsons†
Department of Geography
University of California, Berkeley
Berkeley, California

K. David Patterson†
Department of History
University of North Carolina
Charlotte, North Carolina

Peter L. Pellett
Department of Nutrition
School of Public Health and Health Sciences
University of Massachusetts
Amherst, Massachusetts

Judy Perkin
Department of Health Sciences
Santa Fe Community College
Gainesville, Florida

David M. Peterson
U.S. Department of Agriculture
Agricultural Research Service – Midwest Area
Cereal Crops Research Unit
Madison, Wisconsin

Julia Peterson
School of Nutritional Science and Policy
Tufts University
Medford, Massachusetts

Jeffrey M. Pilcher
Department of History
The Citadel
Charleston, South Carolina

Nancy J. Pollock
Department of Anthropology
Victoria University of Wellington
Wellington, New Zealand

Ananda S. Prasad
Harper-Grace Hospital
Wayne State University School of Medicine
Detroit, Michigan

Heather Munro Prescott
Department of History
Central Connecticut State University
New Britain, Connecticut

Sara A. Quandt
Department of Public Health Sciences
Bowman Gray School of Medicine
Wake Forest University
Winston-Salem, North Carolina

Ted A. Rathbun
Department of Anthropology
University of South Carolina
Columbia, South Carolina

Linda J. Reed
Archaeologist
Burns Paiute Tribe
Burns, Oregon

Elizabeth J. Reitz
Museum of Natural History
University of Georgia
Athens, Georgia

Daphne A. Roe†
Division of Nutritional Sciences
Cornell University
Ithaca, New York

Delphine Roger
Department of History
Université de Paris
Saint-Denis, France

Paul Rozin
Department of Psychology
University of Pennsylvania
Philadelphia, Pennsylvania

Françoise Sabban
École des Hautes Études en Sciences Sociales
Paris, France

Joy B. Sander
Department of Anthropology
University of Colorado
Boulder, Colorado

Ritu Sandhu
Frances Stern Nutrition Center
New England Medical Center
Boston, Massachusetts

Nevin S. Scrimshaw
Food and Nutrition Programme for Human and Social
Development
United Nations University
Boston, Massachusetts

Stephen Seely
Department of Cardiology
The Royal Infirmary
University of Manchester
Manchester, England

Donald T. Simeon
Commonwealth Caribbean Medical Research Council
Port-of-Spain, Trinidad

Wayne R. Snodgrass
Department of Pediatrics
University of Texas Medical Branch
Galveston, Texas

Kristin D. Sobolik
Department of Anthropology
University of Maine
Orono, Maine

Thomas Sorosiak
American Culture Studies
Bowling Green State University
Bowling Green, Ohio

Colin Spencer
Freelance writer and researcher
Suffolk, England

Herta Spencer
Metabolic Research
Veterans Administration Hospital
Hines, Illinois

William J. Stadelman
Department of Food Science
Purdue University
West Lafayette, Indiana

Elizabeth A. Stephens
Department of Anthropology
University of Arizona
Tucson, Arizona

Patricia Stuart-Macadam
Department of Anthropology
University of Toronto
Toronto, Canada

John C. Super
Department of History
West Virginia University
Morgantown, West Virginia

H. Micheal Tarver
Department of History
McNeese State University
Lake Charles, Louisiana

Gretchen Theobald
Department of Anthropology
National Museum of Natural History
Smithsonian Institution
Washington, D.C.

Myrtle Thierry-Palmer
Department of Biochemistry
Morehouse School of Medicine
Atlanta, Georgia

Joyce Toomrey
Davis Center for Russian Studies
Harvard University
Cambridge, Massachusetts

Steven C. Topik
Department of History
University of California, Irvine
Irvine, California

Luis A. Vargas
Instituto de Investigaciones Antropológicas
Universidad Nacional Autónoma de México
Mexico City, Mexico

Keith Vernon
Department of Historical and Critical Studies
University of Central Lancashire
Preston, England

Terrence W. Walters
The Cucurbit Network
P.O. Box 560483
Miami, Florida

Margaret J. Weinberger
American Culture Studies
Bowling Green State University
Bowling Green, Ohio

John H. Weisburger
American Health Foundation
Valhalla, New York

Wilma Wetterstrom
Harvard Botanical Museum
Cambridge, Massachusetts

Barbara Wheaton
Honorary Curator of the Culinary Collection, Schlesinger
Library
Radcliffe College
Cambridge, Massachusetts

William H. Whitaker
School of Social Work
Marywood University
Scranton, Pennsylvania

James C. Whorton
Department of Medical History and Ethics
School of Medicine
University of Washington
Seattle, Washington

Christine S. Wilson, Editor
Ecology of Food and Nutrition
Annapolis, Maryland

Thomas Wilson
Department of Epidemiology
Anthem Blue Cross/Blue Shield
Cincinnati, Ohio

Elizabeth S. Wing
The Florida State Museum
University of Florida
Gainesville, Florida

George Wolf
Department of Nutritional Sciences
University of California, Berkeley
Berkeley, California

Yiming Xia
Institute of Food and Nutrition
Chinese Academy of Preventive Medicine
Beijing, China

David R. Yesner
Department of Anthropology
University of Alaska
Anchorage, Alaska



PREFACE

This work together with its predecessor, *The Cambridge World History of Human Disease*, represents an effort to encapsulate much of what is known about human health as a new millennium begins. As such the volumes should prove important to researchers of the future just as today many investigators find important August Hirsch's three-volume *Handbook of Geographical and Historical Pathology* that was translated and published in London during the years 1883 to 1886. We hope, however, that in light of the accelerating academic and public interest in the things we eat and drink and what they do for us (or to us) the present work will also find an appreciative audience here and now. It is, to our knowledge, the first effort to draw together on a global scale the work of scholars on both nutritional and food-related subjects and to endow the whole with a strong interdisciplinary foundation that utilizes a great number of approaches to food questions ranging from the anthropological to the zoological.

Many of these questions are policy-related and look at real and pressing problems such as poor nutrition among the young of the developing world, food additives and biotechnology in the developed world, and food entitlements in both worlds. Many others, however, are dedicated to determining what our ancestors ate in Paleolithic times; the changes – both dietary and physiological – brought on by the various Neolithic Revolutions and the domesticated plants and animals they nurtured; and what products of those Revolutions have been consumed in the various corners of the globe ever since.

Another broad set of questions employs nutritional science to evaluate the quality of diets, both past and present, and to indicate the diseases, defects,

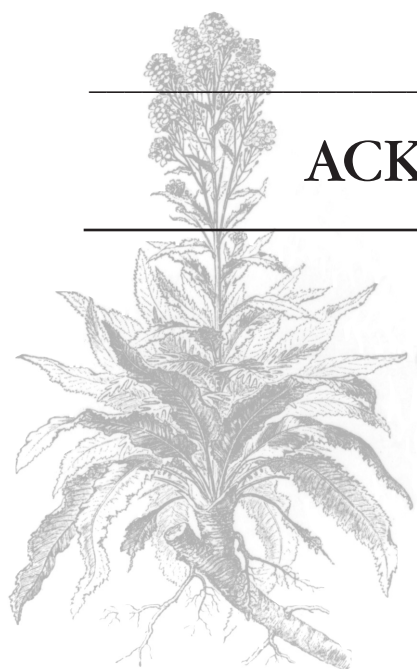
and disorders that can and have developed when that quality is low. A final focus is on the numerous psychological, cultural, and genetic reasons individual humans as well as societies embrace some foods and beverages yet reject others that counterparts find eminently satisfactory.

More implicit than explicit are two threads that loosely stitch together the essays that follow. One is the thread of food globalization, although by globalization we do not mean the current concern about world “burgerization” or even the process of menu homogenization – the latter which began in the aftermath of World War II and has been accelerating ever since. Rather, we mean the inexorable process of food globalization that started some eight to ten thousand or more years ago with the domestication of plants and animals. Although this first occurred in places such as Mesopotamia, the Indus Valley, Egypt, and China in the Old World and Peru and MesoAmerica in the New World, and the fruits of domestication were at first shared mostly by propinquous peoples, those fruits sooner or later spread out over the entire globe. They were carried by restless folk – initially by explorers, pioneers, and the like and then by merchants, missionaries, marauders, and mariners who followed on their heels.

The second thread has to do with the taxes that technological advances generally levy on human nutrition. Many believe that this process began in the distant past as hunter-gatherers devised tools which enabled them to become so efficient in hunting large animals that their growth in numbers and a decline in the amount of prey seriously jeopardized their food supply. This forced another technological advance – the greatest of all – which was the invention of agri-

culture. But with sedentism came even more population growth, circumscribed diets, and pathogenic prosperity. Such problems continue to curse many in the developing world today where the response to any technologically driven increase in the food supply has generally been an increase in the number

who share that food. Meanwhile, in the developed world the response to ever more abundance - much of it in the form of new calorie-dense foods - has been a different kind of population growth as its citizens individually grow ever larger in an outward direction.



ACKNOWLEDGMENTS

This work is the offspring of the Cambridge History and Culture of Food and Nutrition project which, in turn, was the progeny of the Cambridge History of Human Disease project. As the latter effort neared completion in 1990, the former was launched. We are enormously grateful to the National Library of Medicine for providing one grant (1 RO1 LMO5320-01) to begin the effort and another (1RO1 LMOC574-01) to see it completed. In between, funding was provided by a Bowling Green State University History Department challenge grant from the State of Ohio, by grants from Cambridge University Press, and by various university budgets. In addition, Department of History chairpersons Gary Hess, Fujiya Kawashima, and Donald Nieman each ensured that the project was always staffed by graduate students who checked sources, entered manuscripts into the computer, and performed countless other duties. That some four generations of students worked on the project can be seen in the list of assistant editors whose names appear opposite the title page. We are indebted to all of them. In addition, we wish to thank the History Department secretaries Connie Willis and Judy Gilbert for the countless hours of their labor given cheerfully and generously.

Frank Smith, our editor at Cambridge University Press, was instrumental in launching this project. He collaborated in its conceptualization, worked with us in assembling the Board Members, encouraged us at every stage of the effort for a full decade, procured last-minute permissions, and took over the task of securing the decorative art. He has been a true partner in the project.

Rachael Graham, who served as executive editor of the disease project, retired with husband Jim

shortly after the food and nutrition project began. She was succeeded by Mike Tarver (also a veteran of the earlier effort) – a computer whiz who propelled all of us in the direction of computer literacy while overseeing the daily office procedures of entering manuscripts into the computer and checking sources. But then he, too, left us (for “a real job” as we heard one of our assistant editors mutter under his breath), and Stephen Beck took over.

Steve was born for the job, and many of our authors, including myself, are in his debt for sentences that are now clearer and sharper (and in some instances far more coherent) than they were before. His grasp of the English language is also matched by a fine eye for details and a much appreciated ability to keep track of those details over months and even years. This is particularly apparent in Part VIII where he maintained an ever-growing list of the names of plant foods, both common and scientific, which prevented much repetition, not to mention duplication. In addition he and Coneè supervised the activities of the assistant editors on the project and finally, on numerous occasions (such as the pellagra essay, and some of the Part VIII entries), Steve also took on the tasks of researcher and author.

None of this is to imply that others had no hand in the editorial and production process. Kathi Unger and Steve Shimer compiled the very large indexes on the other end of this work both swiftly and competently. Kathie Kounouklos produced the handsome page layouts, and Mary Jo Rhodes steadfastly checked the manuscript and proof from beginning to end. Claire McKean and Phyllis Berk under the direction of Françoise Bartlett at G&H SOHO did a wonderful job of imposing order on a manuscript whose very

size doubtlessly made it intimidating. In so doing they untangled and straightened out much that we thought we had untangled and straightened out. We are very grateful to them and to Cathy Felgar of Cambridge University Press who was ultimately in charge of the work's production.

However, despite the editorial exertions here and in New York, in the final analysis these volumes belong to its authors – in this case around 160 of them – of which fully a quarter represent some 15 countries in addition to the United States. We thank each one for what has become a magnificent collective achievement and apologize for the “extra” years that it took to get the work into print. As we have explained (or confessed) to the many who have inquired, the project just became so large that every one of its stages required considerably more time than estimates based on the earlier disease project allowed for.

The nearly 40 members of our Board of Editors have been a vital part of the project. Most of the suggestions for authors came from them, and every essay was read by them for scientific and historical accuracy. Especially active in this latter capacity were Thomas G. Benedek, Kenneth J. Carpenter, Alfred W. Crosby, Frederick L. Dunn, Daniel W. Gade, Jerome S. Handler, Clark Spencer Larsen, Leslie Sue Lieberman, Ellen Messer, Marion Nestle, James L. Newman, K. David Patterson, and Jeffrey M. Pilcher. In addition, Frederick J. Simoons not only read a number of essays with his characteristically painstaking care but made numerous and invariably sterling suggestions for authors that lightened our load considerably.

It should be noted, however, that we did not always take the suggestions of Board Members. In

some cases this was because reports disagreed. In others it was because authors made good cases against proposed revisions. And at times it was because, rightly or wrongly, we decided that suggested revisions were not necessary. But this means only we and the authors bear final responsibility for the essays embraced by these volumes. In no way does it detract from the great wisdom (not to mention the great amount of time) so graciously donated by the members of our Board of Editors.

One of these individuals, K. David Patterson, died suddenly and prematurely during the latter stages of the project. Dave was a good friend as well as colleague who served both of our Cambridge projects as an energetic board member and enthusiastic author. He did everything from collaborating in their conceptualization to accepting huge writing assignments for each to the researching and writing of essays whose authors failed to deliver them. Yet, despite taking on so much, everything that he did reflected scholarship of the highest possible quality. Dave truly was a scholar in the best sense of the word, and we greatly miss him.

This work is dedicated to Dave and to four other individuals who died during this project. We did not personally know Norman Kretchmer, one of the board members, or authors Richard P. Palmieri, James J. Parsons, and Daphne A. Roc, but we knew them by their work and reputations. We know that they will be sorely missed in, among other fields, those of the Nutritional Sciences and Geography.

KENNETH F. KIPLE
KRIEMHILD CONEË ORNELAS



ABOUT THESE VOLUMES AND HOW TO USE THEM

The history of the world's major staple foods – both animal and vegetable – along with supplementary foods that loom large in such diets – are dealt with at length in Part II. Shorter entries of these same staple plant foods also appear in Part VIII along with some 1,000 of the world's other fruits and vegetables.

These volumes were never intended to comprise an encyclopedia but rather to be a collection of original essays. Therefore, the chapters are far from uniform, and at times there is overlap between them, which was encouraged so that each essay could stand alone without cross-referencing. Bibliographies range from gargantuan to gaunt; most (but not all) authors employ in-text citations, and a few use end notes as well. Such disproportion was not considered a problem. Our authors represent many disciplines, each with his or her own manner of delivering scholarship, and, consequently, they were given considerably latitude in their style of presentation.

The table of contents, of course, is one way of navigating this work, and to help in its use the entries have been arranged in alphabetical order wherever it made sense to do so. But although the table of contents will direct readers to a major entry such as “Wheat,” it is in the index where every mention of wheat in both volumes is indicated and the grain's geographic spread and historical uses can be discerned.

The capability for tracing foods longitudinally came with the preparation of a special subject index that was also geographically focused; hence, a food like manioc can be found both by subject and under the various areas of the world where it is, or has been, important, such as Brazil, the Caribbean, and West Africa. In addition, there is an index of all names mentioned in the text.

The entries in Part VIII are also represented in the subject index. This is the only part of the work that is cross-referenced, and it contains numerous synonyms that should prove useful to researchers and intriguing to linguists, including those no longer in use such as “colewort” (cabbage) or “pompion” (pumpkin). The synonyms appear twice: once in alphabetical order to direct readers to (in the case of the examples just used) “Cabbage” or “Pumpkin” and then again in a synonyms list at the end of these entries. Moreover, the Part VIII articles send the reader to related entries in that section and also to chapters in Parts II and III (in the case of cabbage or pumpkin to “Cruciferous and Green Leafy Vegetables” and “Squash,” respectively).

We should note for the record that any discussion in these volumes of a foodstuff (or a nonfood for that matter) which might be consumed should not be taken as a recommendation that it be consumed, or that it is safe to consume. The new (and sometimes unfamiliar) foods that our authors state are edible doubtlessly are. But as many of the chapters that follow indicate, foods, especially plant foods, are generally toxic to some degree to help protect against predators.

A good example is bitter manioc whose tubers contain prussic acid – a poison that must be removed before consumption. But toxicity can vary with plant parts. Thus, the celery-like stalks of rhubarb are edible after cooking, but the rhizomes and leaves of the plant are so loaded with oxalic acid as to be deadly. Even the common white potatoes that many of us eat may contain solanine in their skins that can make a person sick.

Some plant foods – kava, khat, and coca, for example – are frequently illegal to possess; and still other

potential foods, such as the numerous fungi and myriad other wild plants, require specialized knowledge to sort out the edible from the fatal.

Finally, the chapters on vitamins and minerals discuss the importance of these nutrients to human health both historically and today. In Part VIII “ball-

park” nutrient values are provided for many of the foods listed there. But our authors on the nutrients warn against too much of a good thing, just as our ancient ancestors would have done. Caution and moderation in matters of food were vital to their survival and, by extension, to our presence.



Introduction

We began work on the Cambridge History and Culture of Food and Nutrition Project even as we were still reading the page proofs for *The Cambridge World History of Human Disease*, published in 1993. At some point in that effort we had begun to conceive of continuing our history of human health by moving into food and nutrition – an area that did more than simply focus on the breakdown of that health. For the history of disease we had something of a model provided by August Hirsch in his three-volume *Handbook of Geographical and Historical Pathology* (London, 1883–6). Yet there was no “Handbook of Geographical and Historical Food and Nutrition” to light the way for the present volumes, and thus they would be unique.

Fortunately, there was no lack of expertise available; it came from some 200 authors and board members, representing a score of disciplines ranging from agronomy to zoology. This undertaking, then, like its predecessor, represents a collective interdisciplinary and international effort, aimed in this case at encapsulating what is known of the history of food and nutrition throughout humankind’s stay on the planet. We hope that, together, these volumes on nutrition and the earlier one on disease will provide scholars of the future – as well as those of the present – a glimpse of what is known (and not known) about human health as the twentieth century comes to a close.

Two of our major themes are embedded in the title. *Food*, of course, is central to history; without it, there would be no life and thus no history, and we devote considerable space to providing a history of the most important foodstuffs across the globe. To some extent, these treatments are quantitative, whereas by contrast, *Nutrition* – the body’s need for foods and the

uses it makes of them – has had much to do with shaping the quality of human life. Accordingly, we have placed a considerable array of nutritional topics in longitudinal contexts to illustrate their importance to our past and present and to suggest something of our nutritional future.

The word “Culture,” although not in the book title, was a part of the working title of the project, and certainly the concept of culture permeates the entire work, from the prehistoric culture of our hunting-and-gathering ancestors, through the many different food cultures of the historical era, to modern “food policies,” the prescription and implementation of which are frequently generated by cultural norms. Finally, there is “health,” which appears in none of our titles but is – either explicitly or implicitly – the subject of every chapter that follows and the *raison d’être* for the entire work.

An Overview

Functionally, it seems appropriate to begin this overview of the work with an explanation of the last part first, because we hope that the entries in Part VIII, which identify and sketch out brief histories of vegetable foods mentioned in the text, will constitute an important tool for readers, especially for those interested in the chapters on geographic regions. Moreover, because fruits have seldom been more than seasonal items in the diet, all of these save for a few rare staples are treated in Part VIII. Most readers will need little explanation of foods such as potatoes (also treated in a full chapter) or asparagus but may want to learn more about lesser-known or strictly regional foods such as ackee or zamia (mentioned in the chapters that deal

with the Caribbean area). On the one hand, Part VIII has spared our authors the annoyance of writing textual digressions or footnotes to explain such unfamiliar foods, and on the other, it has provided us with a splendid opportunity to provide more extensive information on the origins and uses of the foods listed. In addition, Part VIII has become the place in the work where synonyms are dealt with, and readers can discover (if they do not already know) that an aubergine is an eggplant, that "swedes" are rutabagas, and that "Bulgar" comes from *bulghur*, which means "bruised grain."

We now move from the end to the beginning of the work, where the chapters of Part I collectively constitute a bioanthropological investigation into the kinds and quantities of foods consumed by early humans, as well as by present-day hunter-gatherers. Humans (in one form or another) have been around for millions of years, but they only invented agriculture and domesticated animals in the past 10,000 years or so, which represents just a tiny fraction of 1 percent of the time humankind has been present on earth. Thus, modern humans must to some extent be a product of what our ancient ancestors ate during their evolutionary journey from scavengers to skilled hunters and from food gatherers to food growers.

The methods for discovering the diet (the foods consumed) and the nutritional status (how the body employed those foods) of our hunting-and-gathering forebears are varied. Archaeological sites have yielded the remains of plants and animals, as well as human coprolites (dried feces), that shed light on the issue of diet, whereas analysis of human remains – bones, teeth, and (on occasion) soft tissue – has helped to illuminate questions of nutrition. In addition, the study of both diet and nutrition among present-day hunter-gatherers has aided in the interpretation of data generated by such archaeological discoveries. The sum of the findings to date seems to suggest that at least in matters of diet and nutrition, our Paleolithic ancestors did quite well for themselves and considerably better than the sedentary folk who followed. In fact, some experts contend that the hunter-gatherers did better than any of their descendants until the late nineteenth and early twentieth centuries.

Part II shifts the focus from foraging to farming and the domestication of plants and animals. The transition from a diet of hunted and collected foods to one based on food production was gradual, yet because its beginnings coincided with the time that many large game animals were disappearing, there is suspicion that necessity, born of an increasing food scarcity, may have been the mother of agricultural invention. But however the development of sedentary agriculture came about, much of the blame for the nutritional deterioration that appears to have accompanied it falls on the production of the so-called superfoods – rice, maize, manioc, and wheat – staples that have sustained great numbers of people but only at a considerable cost in human health, in no small

part because diets that centered too closely on such foods could not provide the range of vitamins, minerals, and whole protein so vital to human health.

Part II is divided into sections, or groups of chapters, most of which consider the history of our most important plant foods under a number of rubrics ranging from "Grains," "Roots, Tubers, and Other Starchy Staples," through "Important Vegetable Supplements," to plants that are used to produce oils and those employed for flavorings. All of the chapters dealing with plants treat questions of where, how, and by whom they were first domesticated, along with their subsequent diffusion around the globe and their present geographic distribution. With domestication, of course, came the dependence of plants on humans along with the reverse, and this phenomenon of "mutualism" is explored in some detail, as are present-day breeding problems and techniques.

The historical importance of the migration of plant foods, although yet to be fully weighed for demographic impact, was vital – although frequently disruptive – for humankind. Wheat, a wild grass that flourished in the wake of retreating glaciers some 12,000 years ago, was (apparently) deliberately planted for the first time in the Middle East about 2,000 years later. By the first century B.C., Rome required some 14 million bushels per year just to feed the people of that city, leading to a program of expansion that turned much of the cultivable land of North Africa into wheatfields for the Romans. Surely, then, Italians had their pastas long before Marco Polo (1254?-1324?), who has been credited with bringing notions of noodles back with him from China. But it was only the arrival of the vitamin C-loaded American tomato that allowed the Italians to concoct the great culinary union of pasta and tomato sauce – one that rendered pasta not only more satisfactory but also more healthy. And, speaking of China, the New World's tomato and its maize, potatoes, sweet potatoes, and peanuts were also finding their respective ways to that ancient land, where in the aftermath of their introduction, truly phenomenal population increases took place.

Migrating American plants, in other words, did much more than just dress up Old World dishes, as tomatoes did pasta. Maize, manioc, sweet potatoes, a new kind of yam, peanuts, and chilli peppers reached the western shores of Africa with the ships of slave traders, who introduced them into that continent to provide food for their human cargoes. Their success exceeded the wildest of expectations, because the new foods not only fed slaves bound for the Americas but helped create future generations of slaves. The American crops triggered an agricultural revolution in Africa, which in greatly expanding both the quantity and quality of its food supply, also produced swelling populations that were drained off to the Americas in order to grow (among other things) sugar and coffee – both migrating plants from the Old World.

In Europe, white potatoes and maize caught on more slowly, but the effect was remarkably similar. Old

World wheat gave back only 5 grains for every 1 planted, whereas maize returned 25 to 100 (a single ear of modern maize yields about 1,000 grains) and, by the middle of the seventeenth century, had become a staple of the peasants of northern Spain, Italy, and to a lesser extent, southern France. From there maize moved into much of the rest of Europe, and by the end of the eighteenth century, such cornmeal mushes (*polenta* in Italy) had spread via the Ottoman Empire into the Balkans and southern Russia.

Meanwhile, over the centuries, the growth of cities and the development of long-distance trade – especially the spice trade – had accelerated the process of exploring the world and globalizing its foods. So, too, had the quest for oils (to be used in cooking, food preservation, and medicines), which had been advanced as coconuts washed up on tropical shores, olive trees spread across the Mediterranean from the Levant to the rim of the Atlantic in Iberia, and sesame became an integral part of the burgeoning civilizations of North Africa and much of Asia.

In the seventeenth century, invasion, famine, and evictions forced Irish peasants to adopt the potato as a means of getting the most nourishment from the least amount of cultivated land, and during the eighteenth century, it was introduced in Germany and France because of the frequent failures of other crops. From there, the plant spread toward the Ural Mountains, where rye had long been the only staple that would ripen during the short, often rainy summers. Potatoes not only did well under such conditions, they provided some four times as many calories per acre as rye and, by the first decades of the nineteenth century, were a crucial dietary element in the survival of large numbers of northern Europeans, just as maize had become indispensable to humans in some of the more southerly regions.

Maize nourished humans indirectly as well. Indeed, with maize available to help feed livestock, it became increasingly possible to carry more animals through the winters and to derive a steady supply of whole protein in the forms of milk, cheese, and eggs, in addition to year-round meat – now available for the many rather than the few. Thus, it has been argued that it is scarcely coincidence that beginning in the eighteenth century, European populations began to grow and, by the nineteenth century, had swollen to the point where, like the unwilling African slaves before them, Europeans began migrating by the millions to the lands whose plants had created the surplus that they themselves represented.

The last section of Part II treats foods from animal sources ranging from game, bison, and fish to the domesticated animals. Its relatively fewer chapters make clear the dependence of all animals, including humans, on the plant world. In fact, to some unmeasurable extent, the plant foods of the world made still another important contribution to the human diet by assisting in the domestication of those animals that – like the dog that preceded them – let themselves be tamed.

The dog seems to have been the first domesticated animal and the only one during the Paleolithic age. The wolf, its progenitor, was a meat eater and a hunter (like humans), and somewhere along the way, humans and dogs seem to have joined forces, even though dogs were sometimes dinner and probably vice versa. But it was during the early days of the Neolithic, as the glaciers receded and the climate softened, that herbivorous animals began to multiply, and in the case of sheep and goats, their growing numbers found easy meals in the grains that humans were raising (or at least had staked out for their own use). Doubtless, it did not take the new farmers long to cease trying to chase the animals away and to begin capturing them instead – at first to use as a source of meat to go with the grain and then, perhaps a bit later, to experiment with the fleece of sheep and the waterproof hair of goats.

There was, however, another motive for capturing animals, which was for use in religious ceremonies involving animal sacrifice. Indeed, it has been argued that wild water buffalo, cattle, camels, and even goats and sheep were initially captured for sacrifice rather than for food.

Either way, a move from capturing animals to domestication and animal husbandry was the next step in the case of those animals that could be domesticated. In southeastern Europe and the Near East (the sites of so much of this early activity), wild goats and sheep may have been the first to experience a radical change of lifestyle – their talent for clearing land of anything edible having been discovered and put to good use by their new masters. Soon, sheep were being herded, with the herdsmen and their flocks spreading out far and wide to introduce still more humans to the mysteries and rewards of domestication.

Wild swine, by contrast, were not ruminant animals and thus were not so readily attracted to the plants in the fields, meaning that as they did not come to humans, humans had to go to them. Wild boars had long been hunted for sacrifice as well as for meat and would certainly have impressed their hunters with their formidable and ferocious nature. Tricky, indeed, must have been the process that brought the domesticated pig to the barnyard by about 7000 to 6000 B.C.

Wild cattle were doubtless drawn to farmers' fields, but in light of what we know about the now-extinct aurochs (the wild ancestor of our modern cattle), the domestication of bovines around 6000 B.C. may have required even more heroic efforts than that of swine. Yet those efforts were certainly worth it, for in addition to the meat and milk and hides cattle provided, the ox was put to work along with sheep and goats as still another hand in the agricultural process – stomping seeds into the soil, threshing grain, and pulling carts, wagons, and (later on) the plow.

The last of today's most important animals to be domesticated was the chicken, first used for sacrifice and then for fighting before it and its eggs became food. The domesticated variety of this jungle bird was

present in North China around 3000 B.C.; however, because the modern chicken is descended from both Southeast Asian and Indian wildfowl, the question of the original site of domestication has yet to be resolved. The wildfowl were attracted to human-grown grain and captured, as was the pigeon (which, until recently, played a far more important role in the human diet than the chicken). Ducks, geese, and other fowl were also most likely captivated by – and captured because of – the burgeoning plant-food products of the Neolithic. In other parts of the world, aquatic animals, along with the camel, the yak, and the llama and alpaca, were pressed into service by *Homo sapiens*, the “wise man” who had not only scrambled to the top of the food chain but was determinedly extending it.

The chapters of Part III focus on the most important beverages humans have consumed as accompaniment to those foods that have preoccupied us to this point. One of these, water, is crucial to life itself; another, human breast milk, has – until recently, at least – been vital for the survival of newborns, and thus vital for the continuation of the species. Yet both have also been sources of infection for humans, sometimes fatally so.

Hunter-gatherers, in general, did not stay in one place long enough to foul springs, ponds, rivers, and lakes. But sedentary agriculturalists did, and their own excreta was joined by that of their animals. Wherever settlements arose (in some cases as kernels of cities to come), the danger of waterborne disease multiplied, and water – essential to life – also became life-threatening. One solution that was sensible as well as pleasurable lay in the invention of beverages whose water content was sterilized by the process of fermentation. Indeed, the earliest written records of humankind mention ales made from barley, millet, rice, and other grains, along with toddies concocted from date palms and figs – all of which makes it apparent that the production of alcohol was a serious business from the very beginning of the Old World Neolithic.

It was around 3000 B.C. that grape wine made its appearance, and where there was honey there was also mead. The discovery of spirit distillation to make whiskeys and brandies began some seven to eight hundred years ago, and true beer, the “hopped” successor of ales, was being brewed toward the end of the Middle Ages (about 600 years ago). Clearly, humans long ago were investing much ingenuity in what can only be described as a magnificent effort to avoid waterborne illness.

Milk, one of the bonuses of animal domestication, was also fermented, although not always with desired outcomes. Yet over time, the production of yoghurts, cheeses, and butter became routine, and these foods – with their reduced lactose – were acceptable even among the lactose-intolerant, who constituted most of the world’s population. Where available, milk (especially bovine milk) was a food for the young after

weaning, and during the past few centuries, it has also served as a substitute for human milk for infants, although sometimes with disastrous results. One problem was (and is) that the concentrated nutrient content of bovine milk, as well as human antibodies developed against cow’s-milk protein, make it less than the perfect food, especially for infants. But another was that bovine tuberculosis (scrofula), along with ordinary tuberculosis, raged throughout Europe from the sixteenth to the nineteenth centuries. Wet nurses were another solution for infant feeding, but this practice could be fraught with danger, and artificial feeding, especially in an age with no notions of sterile procedure, caused infants to die in staggering numbers before the days of Joseph Lister and Louis Pasteur.

Boiling water was another method of avoiding the pathogens it contained, and one that, like fermentation, could also produce pleasant beverages in the process. The Chinese, who had used tea since the Han period, embraced that beverage enthusiastically during the Tang dynasty (618–907) and have been avid tea drinkers ever since. The nomads of central Asia also adopted the drink and later introduced it into Russia. Tea use spread to Japan about the sixth century, but it became popular there only about 700 years ago. From Japan, the concoction was introduced into Indonesia, where much later (around 1610) the Dutch discovered it and carried it to Europe. A few decades later, the English were playing a major role in popularizing the beverage, not to mention merchandising it.

Coffee, although it found its way into Europe at about the same time as tea, has a more recent history, which, coffee-lore would have it, began in Ethiopia in the ninth century. By 1500, coffee drinking was widespread throughout the Arab world (where alcohol was forbidden), and with the passing of another couple of centuries, the beverage was enjoying a considerable popularity in Europe. Legend has it that Europeans began to embrace coffee after the Ottoman Turks left some bags of coffee beans behind as they gave up the siege of Vienna in 1683.

These Asian and African contributions to the world’s beverages were joined by cacao from America. Because the Spaniards and the Portuguese were the proprietors of the lands where cacao was grown, they became the first Europeans to enjoy drinking chocolate (which had long been popular among pre-Columbian Mesoamericans). In the early decades of the sixteenth century, the beverage spread through Spain’s empire to Italy and the Netherlands and, around midcentury, reached England and France.

Thus, after millennia of consuming alcoholic beverages to dodge fouled water, people now had (after a century or so of “catching on”) an opportunity for relative sobriety thanks to these three new drinks, which all arrived in Europe at about the same time. But an important ingredient in their acceptance was the sugar that sweetened them. And no wonder that as these beverages gained in popularity, the slave

trade quickened, plantation societies in the Americas flourished, and France in 1763 ceded all of Canada to Britain in order to regain its sugar-rich islands of Martinique and Guadeloupe.

Sugar cultivation and processing, however, added still another alcoholic beverage – rum – to a growing list, and later in the nineteenth century, sugar became the foundation of a burgeoning soft-drink industry. Caffeine was a frequent ingredient in these concoctions, presumably because, in part at least, people had become accustomed to the stimulation that coffee and tea provided. The first manufacturers of Coca-Cola in the United States went even further in the pursuit of stimulation by adding coca – from the cocaine-containing leaves that are chewed in the Andean region of South America. The coca was soon removed from the soft drink and now remains only in the name Coca-Cola, but “cola” continued as an ingredient. In the same way that coca is chewed in South America, in West Africa the wrapping around the kola nut is chewed for its stimulative effect, in this case caused by caffeine. But the extract of the kola nut not only bristles with caffeine, it also packs a heart stimulant, and the combination has proven to be an invigorating ingredient in the carbonated beverage industry.

In East Africa, the leaves of an evergreen shrub called khat are chewed for their stimulating effect and are made into a tealike beverage as well. And finally, there is kava, widely used in the Pacific region and among the most controversial, as well as the most exotic, of the world’s lesser-known drinks – controversial because of alleged narcotic properties and exotic because of its ceremonial use and cultural importance.

In addition to the beverages that humans have invented and imbibed throughout the ages as alternatives to water, many have also clung to their “waters.” Early on, special waters may have come from a spring or some other body of water, perhaps with supposed magical powers, or a good flavor, or simply known to be safe. In more recent centuries, the affluent have journeyed to mineral springs to “take the waters” both inside and outside of their bodies, and mineral water was (and is) also bottled and sold for its allegedly healthy properties. Today, despite (or perhaps because of) the water available to most households in the developed world, people have once more staked out their favorite waters, and for some, bottled waters have replaced those alcoholic beverages that were previously employed to avoid water.

Part IV focuses on the history of the discovery and importance of the chief nutrients, the nutritional deficiency diseases that occur when those nutrients are not forthcoming in adequate amounts, the relationship between modern diets and major chronic diseases, and food-related disorders. Paradoxically, many such illnesses (the nutritional deficiency diseases in particular), although always a potential hazard, may have become prevalent among humans only as a result of the development of sedentary agriculture.

Because such an apparently wide variety of domesticated plant and animal foods emerged from the various Neolithic revolutions, the phenomenon of sedentary agriculture was, at least until recently, commonly regarded as perhaps humankind’s most important step up the ladder of progress. But the findings of bioanthropologists (discussed in Part I) suggest rather that our inclination to think of history teleologically had much to do with such a view and that progress imposes its own penalties (indeed, merely to glance at a newspaper is to appreciate why many have begun to feel that technological advances should carry health-hazard warnings).

As we have already noted, with agriculture and sedentism came diets too closely centered on a single crop, such as wheat in the Old World and maize in the New, and although sedentism (unlike hunting and gathering) encouraged population growth, such growth seems to have been that of a “forced” population with a considerably diminished nutritional status.

And more progress seems inevitably to have created more nutritional difficulties. The navigational and shipbuilding skills that made it possible for the Iberians to seek empires across oceans also created the conditions that kept sailors on a diet almost perfectly devoid of vitamin C, and scurvy began its reign as the scourge of seamen. As maize took root in Europe and Africa as well as in the U.S. South, its new consumers failed to treat it with lime before eating – as the Native Americans, presumably through long experience, had learned to do. The result of maize in inexperienced hands, especially when there was little in the diet to supplement it, was niacin deficiency and the four Ds of pellagra: dermatitis, diarrhea, dementia, and death. With the advent of mechanical rice mills in the latter nineteenth century came widespread thiamine deficiency and beriberi among peoples of rice-eating cultures, because those mills scraped away the thiamine-rich hulls of rice grains with energetic efficiency.

The discovery of vitamins during the first few decades of the twentieth century led to the food “fortification” that put an end to the classic deficiency diseases, at least in the developed world, where they were already in decline. But other health threats quickly took their place. Beginning in the 1950s, surging rates of cancer and heart-related diseases focused suspicion on the environment, not to mention food additives such as monosodium glutamate (MSG), cyclamates, nitrates and nitrites, and saccharin. Also coming under suspicion were plants “engineered” to make them more pest-resistant – which might make them more carcinogenic as well – along with the pesticides and herbicides, regularly applied to farm fields, that can find their way into the human body via plants as well as drinking water.

Domesticated animals, it has turned out, are loaded with antibiotics and potentially artery-clogging fat, along with hormones and steroids that stimulate the

growth of that fat. Eggs have been found to be packed with cholesterol, which has become a terrifying word, and the fats in whole milk and most cheeses are now subjects of considerable concern for those seeking a “heart-healthy” diet. Salt has been implicated in the etiology of hypertension, sugar in that of heart disease, saturated fats in both cancer and heart disease, and a lack of calcium in osteoporosis. No wonder that despite their increasing longevity, many people in the developed world have become abruptly and acutely anxious about what they do and do not put in their mouths.

Ironically, however, the majority of the world’s people would probably be willing to live with some of these perils if they could share in such bounty. Obesity, anorexia, and chronic disease might be considered tolerable (and preferable) risks in the face of infection stalking their infants (as mothers often must mix formulas with foul water); protein-energy malnutrition attacking the newly weaned; iodine deficiency (along with other mineral and vitamin deficiencies) affecting hundreds of millions of children and adults wherever foods are not fortified; and undernutrition and starvation. All are, too frequently, commonplace phenomena.

Nor are developing-world peoples so likely as those in the developed world to survive the nutritional disorders that seem to be legacies of our hunter-gatherer past. Diabetes (which may be the result of a “thrifty” gene for carbohydrate metabolism) is one of these diseases, and hypertension may be another; still others are doubtless concealed among a group of food allergies, sensitivities, and intolerances that have only recently begun to receive the attention they deserve.

On a more pleasant note, the chapters of Part V sketch out the history and culture of food and drink around the world, starting with the beginnings of agriculture in the ancient Near East and North Africa and continuing through those areas of Asia that saw early activity in plant and animal domestication. This discussion is followed by sections on the regions of Europe, the Americas, and sub-Saharan Africa and Oceania.

Section B of Part V takes up the history of food and drink in South Asia and the Middle East, Southeast Asia, and East Asia in five chapters. One of these treats the Middle East and South Asia together because of the powerful culinary influence of Islam in the latter region, although this is not to say that Greek, Persian, Aryan, and central Asian influences had not found their way into South Asia for millennia prior to the Arab arrival.

Nor is it to say that South Asia was without its own venerable food traditions. After all, many of the world’s food plants sprang from the Indus Valley, and it was in the vastness of the Asian tropics and subtropics that most of the world’s fruits originated, and most of its spices. The area is also home to one of our “superfoods,” rice, which ties together the cuisines of much of the southern part of the continent, whereas millet and (later) wheat were the staples of the north-

ern tier. Asia was also the mother of two more plants that had much to do with transforming human history. From Southeast Asia came the sugarcane that would later so traumatize Africa, Europe, and the Americas; from eastern Asia came the evergreen shrub whose leaves are brewed to make tea.

Rice may have been cultivated as many as 7,000 years ago in China, in India, and in Southeast Asia; the wild plant is still found in these areas today. But it was likely from the Yangtze Delta in China that the techniques of rice cultivation radiated outward toward Korea and then, some 2,500 years ago, to Japan. The soybean and tea also diffused from China to these Asian outposts, all of which stamped some similarities on the cuisines of southern China, Japan, and Korea. Northern China, however, also made the contribution of noodles, and all these cuisines were enriched considerably by the arrival of American plants such as sweet potatoes, tomatoes, chillies, and peanuts – initially brought by Portuguese ships between the sixteenth century (China) and the eighteenth century (Japan).

Also characteristic of the diets of East Asians was the lack of dairy products as sources of calcium. Interestingly, the central Asian nomads (who harassed the northern Chinese for millennia and ruled them when they were not harassing them) used milk; they even made a fermented beverage called *kumiss* from the milk of their mares. But milk did not catch on in China and thus was not diffused elsewhere in East Asia. In India, however, other wanderers – the Aryan pastoralists – introduced dairy products close to 4,000 years ago. There, dairy foods did catch on, although mostly in forms that were physically acceptable to those who were lactose-intolerant – a condition widespread among most Asian populations.

Given the greater sizes of Sections C (Europe) and D (the Americas) in Part V, readers may object to what clearly seems to be something of a Western bias in a work that purports to be global in scope. But it is the case that foods and foodways of the West have been more systematically studied than those of other parts of the world, and thus there are considerably more scholars to make their expertise available. In most instances, the authors of the regional essays in both these sections begin with the prehistoric period, take the reader through the Neolithic Revolution in the specific geographic area, and focus on subsequent changes in foodways wrought by climate and cultural contacts, along with the introduction of new foods. At first, the latter involved a flow of fruits and vegetables from the Middle and Near East into Europe, and an early spice trade that brought all sorts of Asian, African, and Near Eastern goods to the western end of the Mediterranean. The expansion of Rome continued the dispersal of these foods and spices throughout Europe.

Needless to say, the plant and animal exchanges between the various countries of the Old World and

the lands of the New World following 1492 are dealt with in considerable detail because those exchanges so profoundly affected the food (and demographic) history of all the areas concerned. Of course, maize, manioc, sweet potatoes and white potatoes, peanuts, tomatoes, chillies, and a variety of beans sustained the American populations that had domesticated and diffused them for a few thousand years in their own Neolithic Revolution before the Europeans arrived. But the American diets were lacking in animal protein. What was available came (depending on location) from game, guinea pigs, seafoods, insects, dogs, and turkeys. That the American Indians did not domesticate more animals – or milk those animals (such as the llama) that they did domesticate – remains something of a mystery. Less of a mystery is the fate of the Native Americans, many of whom died in a holocaust of disease inadvertently unleashed on them by the Europeans. And as the new land became depopulated of humans, it began to fill up again with horses, cattle, sheep, hogs, and other Old World animals.

Certainly, the addition of Old World animal foods to the plants of the New World made for a happy union, and as the authors of the various regional entries approach the present – as they reach the 1960s, in fact – an important theme that emerges in their chapters is the fading of distinctive regional cuisines in the face of considerable food globalization. The cuisine of the developed world, in particular, is becoming homogenized, with even natives of the Pacific, Arctic, and Subarctic regions consuming more in the way of the kinds of prepared foods that are eaten by everybody else in the West, unfortunately to their detriment.

Section E treats the foodways of Africa south of the Sahara, the Pacific Islands, and Australia and New Zealand in three chapters that conclude a global tour of the history and culture of food and drink. Although at first glance it might seem that these last three disparate areas of the planet historically have had nothing in common from a nutritional viewpoint, they do, in fact, share one feature, which has been something of a poverty of food plants and animals.

In Africa, much of this poverty has been the result of rainfall, which depending on location, has generally been too little or too much. Famine results from the former, whereas leached and consequently nitrogen- and calcium-poor soils are products of the latter, with the plants these areas do sustain also deficient in important nutrients. Moreover, 40 inches or more of rainfall favors proliferation of the tsetse fly, and the deadly trypanosomes carried by this insect have made it impossible to keep livestock animals in many parts of the continent. But even where such animals can be raised, the impoverished plants they graze on render them inferior in size, as well as inferior in the quality of their meat and milk, to counterparts elsewhere in the world. As in the Americas, then, animal protein was not prominent in most African diets after the advent of sedentism.

But unlike the Americas, Africa was not blessed with vegetable foods, either. Millets, yams, and a kind of African rice were the staple crops that emerged from the Neolithic to sustain populations, and people became more numerous in the wake of the arrival of better-yielding yams from across the Indian Ocean. But it was only with the appearance of the maize, peanuts, sweet potatoes, American yams, manioc, and chillies brought by the slave traders that African populations began to experience the substantial growth that we still witness today.

Starting some 30,000 to 40,000 years ago, waves of Pacific pioneers spread out from Southeast Asia to occupy the islands of Polynesia, Melanesia, and Micronesia. They lived a kind of fisher-hunter-gatherer existence based on a variety of fish, birds, and reptiles, along with the roots of ferns and other wild vegetable foods. But a late wave of immigrants, who sailed out from Southeast Asia to the Pacific Basin Islands about 6,000 years ago, thoughtfully brought with them some of the products of the Old World Neolithic in the form of pigs, dogs, chickens, and root crops like the yam and taro. And somehow, an American plant – the sweet potato – much later also found its way to many of these islands.

In a very real sense, then, the Neolithic Revolution was imported to the islands. Doubtless it spread slowly, but by the time the ships of Captain James Cook sailed into the Pacific, all islands populated by humans were also home to hogs, dogs, and fowl – and this included even the extraordinarily isolated Hawaiian Islands. Yet, as with the indigenous populations of the Americas, those of the Pacific had little time to enjoy any plant and animal gifts the Europeans brought to them. Instead, they began to die from imported diseases, which greatly thinned their numbers.

The story of Australia and New Zealand differs substantially from that of Africa and the Pacific Islands in that both the Australian Aborigines and (to a lesser extent) the New Zealand Maori were still hunter-gatherers when the Europeans first reached them. They had no pigs or fowl nor planted yams or taro, although they did have a medium-sized domesticated dog and sweet potatoes.

In New Zealand, there were no land mammals prior to human occupation, but there were giant flightless birds and numerous reptiles. The Maori arrived after pigs and taro had reached Polynesia, but at some point (either along the way to New Zealand or after their arrival) they lost their pigs, and the soil and climate of New Zealand did not lend themselves to growing much in the way of taro. Like their Australian counterparts, they had retained their dogs, which they used on occasion for food, and the sweet potato was their most important crop.

Thus, despite their dogs and some farming efforts, the Aborigines and the Maori depended heavily on hunting-and-gathering activities until the Europeans arrived to introduce new plant and animal species.

Unfortunately, as in the Americas and elsewhere in the Pacific, they also introduced new pathogens and, consequently, demographic disaster.

Following this global excursion, Part V closes with a discussion of the growing field of culinary history, which is now especially vigorous in the United States and Europe but promises in the near future to be a feast that scholars the world over will partake of and participate in.

Part VI is devoted to food- and nutrition-related subjects that are of both contemporary and historical interest. Among these are some examples of the startling ability of humans to adapt to unique nutritional environments, including the singular regimen of the Inuit, whose fat-laden traditional diet would seem to have been so perfectly calculated to plug up arteries that one might wonder why these people are still around to study. Other chapters take up questions regarding the nutritional needs (and entitlements) of special age, economic, and ethnic groups. They show how these needs frequently go unmet because of cultural and economic circumstances and point out some of the costs of maternal and child undernutrition that are now undergoing close scrutiny, such as mental decrement. In this vein, food prejudices and taboos are also discussed; many such attitudes can bring about serious nutritional problems for women and children, even though childbearing is fundamentally a nutritional task and growing from infancy to adulthood a nutritional feat.

A discussion of the political, economic, and biological causes and ramifications of famine leads naturally to another very large question treated in the first two chapters of Part VI. The importance of nutrition in humankind's demographic history has been a matter of some considerable debate since Thomas McKeown published *The Modern Rise of Population* in 1976. In that work, McKeown attempted to explain how it happened that sometime in the eighteenth century if not before, the English (and by extension the Europeans) managed to begin extricating themselves from seemingly endless cycles of population growth followed by plunges into demographic stagnation. He eliminated possibilities such as advances in medicine and sanitation, along with epidemiological factors such as disease abatement or mutation, and settled on improved nutrition as the single most important cause. Needless to say, many have bristled at such a high-handed dismissal of these other possibilities, and our chapters continue the debate with somewhat opposing views.

Not entirely unrelated is a discussion of height and nutrition, with the former serving as proxy for the latter. Clearly, whether or not improving nutrition was the root cause of population growth, it most certainly seems to have played an important role in human growth and, not incidentally, in helping at least those living in the West to once again approach the stature of their Paleolithic ancestors. Moreover, it is the case that no matter what position one holds with respect

to the demographic impact of nutrition, there is agreement that nutrition and disease cannot be neatly separated, and indeed, our chapter on synergy describes how the two interact.

Cultural and psychological aspects of food are the focus of a group of chapters that examines why people eat some foods but not others and how such food choices have considerable social and cultural resonance. Food choices of the moment frequently enter the arena of food fads, and one of our chapters explores the myriad reasons why foods can suddenly become trends, but generally trends with little staying power.

The controversial nature of vegetarianism – a nutritional issue always able to trigger a crossfire of debate – is acknowledged in our pages by two chapters with differing views on the subject. For some, the practice falls under the rubric of food as medicine. Then there are those convinced of the aphrodisiacal benefits of vegetarianism – that the avoidance of animal foods positively influences their sexual drive and performance. For many, vegetarianism stems from religious conviction; others simply feel it is wrong to consume the flesh of living creatures, whereas still others think it downright dangerous. Clearly, the phrase “we are what we eat” must be taken in a number of different ways.

The closing chapters of Part VI address the various ways that humans and the societies they construct have embraced particular foods or groups of foods in an effort to manipulate their own health and well-being as well as that of others. Certain foods, for example, have been regarded by individuals as aphrodisiacs and anaphrodisiacs and consumed in frequently heroic efforts to regulate sexual desires. Or again, some – mostly plant – foods have been employed for medicinal reasons, with many, such as garlic, viewed as medical panaceas.

Part VII scrutinizes mostly contemporary food-related policy questions that promise to be with us for some time to come, although it begins with a chapter on nutrition and the state showing how European governments came to regard well-nourished populations as important to national security and military might. Other discussions that follow treat the myriad methodological (not to mention biological) problems associated with determining the individual's optimal daily need for each of the chief nutrients; food labeling, which when done fairly and honestly can aid the individual in selecting the appropriate mix of these nutrients; and the dubious ability of nonfoods to supplement the diet.

As one might expect, food safety, food biotechnology, and the politics of such issues are of considerable concern, and – it almost goes without saying – politics and safety have the potential at any given time for being at odds with one another. The juxtaposition is hardly a new one, with monopoly and competitive capital on the one hand and the public interest on the other. The two may or may not be in opposition, but the stakes are enormous, as will readily be seen.

First there is the problem of safety, created by a loss of genetic diversity. Because all crops evolved from wild species, this means that in Darwinian terms, that the latter possessed sufficient adaptability to survive over considerable periods of time. But with domestication and breeding has come genetic erosion and a loss of this adaptability – even the loss of wild progenitors – so that if today many crops were suddenly not planted, they would simply disappear. And although this possibility is not so alarming – after all, everyone is not going to cease planting wheat, or rice, or maize – the genetic sameness of the wheat or the maize or the rice that is planted (the result of a loss of genetic material) has been of some considerable concern because of the essentially incalculable risk that some newly mutated plant plague might arise to inflict serious damage on a sizable fraction of the world's food supply.

There is another problem connected with the loss of genetic material. It is less potentially calamitous but is one that observers nevertheless find disturbing, especially in the long term. The problem is that many crops have been rendered less able to fend off their traditional parasites (in part because of breeding that reduces a plant's ability to produce the naturally occurring toxicants that defend against predators) and thus have become increasingly dependent on pesticides that can and do find their way into our food and water supplies.

Genetic engineering, however, promises to at least reduce the problem of chemical pollution by revitalizing the ability of crops to defend themselves – as, for example, in the crossing of potatoes with carnivorous plants so that insects landing on them will die immediately. But the encouragement of such defense mechanisms in plants has prompted the worry that because humans are, after all, parasites as far as the plant is concerned, resistance genes might transform crops into less healthy or even unhealthy food, perhaps (as mentioned before) even carcinogenic at some unacceptable level. And, of course, genetic engineering has also raised the specter of scientists accidentally (or deliberately) engineering and then unleashing self-propagating microorganisms into the biosphere, with disastrous epidemiological and ecological effect.

Clearly, biotechnology, plant breeding, plant molecular and cellular biology, and the pesticide industry all have their perils as well as their promise, and some of these dangers are spelled out in a chapter on toxins in foods. But in addition, as a chapter on substitute foods shows, although these substitutes may have been developed to help us escape the tyranny of sugars and fats, they are not without their own risks. Nor, for that matter, are some food additives. Although most seem safe, preservatives such as nitrates and nitrites, flavor enhancers like MSG, and coloring agents such as tartrazine are worrisome to many.

As our authors make clear, however, we may have more to fear from the naturally occurring toxins that the so-called natural foods employ to defend them-

selves against predators than from the benefits of science and technology. Celery, for example, produces psoralins (which are mutagenic carcinogens); spinach contains oxalic acid that builds kidney stones and interferes with the body's absorption of calcium; lima beans have cyanide; and the solanine in the skins of greenish-appearing potatoes is a poisonous alkaloid.

From biological and chemical questions, we move to other problems of a political and economic nature concerning what foods are produced, what quantities are produced, what the quality is of these foods, and what their allocation is. In the United States (and practically everywhere else) many of the answers to such questions are shaped and mediated by lobbying groups, whose interests are special and not necessarily those of the public. Yet if Americans sometimes have difficulty in getting the truth about the foods they eat, at least they get the foods. There is some general if uneasy agreement in America and most of the developed world that everyone is entitled to food as a basic right and that government programs – subsidies, food stamps, and the like – ought to ensure that right. But such is not the situation in much of the developing world, where food too frequently bypasses the poor and the powerless. And as the author of the chapter on food subsidies and interventions makes evident, too often women and children are among the poor and the powerless.

To end on a lighter note, the last chapter in Part VII takes us full circle by examining the current and fascinating issue of the importance of Paleolithic nutrition to humans entering the twenty-first century.

We close this introduction on a mixed note of optimism and pessimism. The incorporation of dwarfing genes into modern plant varieties was responsible for the sensationally high-yielding wheat and rice varieties that took hold in developing countries in the 1960s, giving rise to what we call the "Green Revolution," which was supposed to end world hunger and help most of the countries of the world produce food surpluses. But the Green Revolution also supported a tremendous explosion of populations in those countries it revolutionized, bringing them face to face with the Malthusian poles of food supply and population.

Moreover, the new plants were heavily dependent on the petrochemical industry for fertilizers, so that in the 1970s, when oil prices soared, so did the price of fertilizers, with the result that poorer farmers, who previously had at least eked out a living from the land, were now driven from it. Moreover, the new dwarfed and semidwarfed rice and wheat plants carried the same genes, meaning that much of the world's food supply was now at the mercy of new, or newly mutated, plant pathogens. To make matters worse, the plants seemed even less able to defend themselves against existing pathogens. Here, the answer seemed to be a still more lavish use of pesticides (against which bitter assaults were launched by environmentalists) even as more developing-world farmers were

being driven out of business by increasing costs, and thousands upon thousands of people were starving to death each year. Indeed, by the 1980s, every country revolutionized by the Green Revolution was once again an importer of those staple foods they had expected to produce in abundance.

Obviously, from both a social and political-economic as well as a biological viewpoint, ecologies had not only failed to mesh, they had seriously unraveled. However, as our earlier chapters on rice and wheat point out, new varieties from plant breeders contain variations in genes that make them less susceptible to widespread disease damage, and genetic engineering efforts are under way to produce other varieties that will be less dependent on fertilizers and pesticides.

Meanwhile, as others of our authors point out, foods such as amaranth, sweet potatoes, manioc, and taro, if given just some of the attention that rice and wheat have received, could help considerably to expand the world's food supply. But here again, we

teeter on the edge of matters that are as much cultural, social, economic, and political in nature as they are ecological and biological. And such matters will doubtless affect the acceptance of new crops of nutritional importance.

As we begin a sorely needed second phase of the Green Revolution, observers have expressed the hope that we have learned from the mistakes of the first phase. But of course, we could call the first flowering of the Neolithic Revolution (some 10,000 years ago) the first phase and ponder what has been learned since then, which – in a nutshell – is that every important agricultural breakthrough thus far has, at least temporarily, produced unhappy health consequences for those caught up in it, and overall agricultural advancement has resulted in growing populations and severe stress on the biosphere. As we enter the twenty-first century, we might hope to finally learn from our mistakes.

The Editors



PART I

Determining What Our Ancestors Ate

About 10,000 years ago, humans started changing the way they made a living as they began what would be a lengthy transition from foraging to farming. This transformation, known as the Neolithic Revolution, was actually comprised of many revolutions, taking place in different times and places, that are often viewed collectively as the greatest of all human strides taken in the direction of progress. But such progress did not mean better health. On the contrary, as the following chapters indicate, hunter-gatherers were, on the whole, considerably better nourished and much less troubled with illnesses than their farmer descendants. Because hunter-gatherers were mobile by necessity, living in bands of no more than 100 individuals they were not capable of supporting the kinds of ailments that flourished as crowd diseases later on. Nor, as a rule, did they pause in one spot long enough to foul their water supply or let their wastes accumulate to attract disease vectors – insects, rodents, and the like. In addition, they possessed no domesticated animals (save the dog late in the Paleolithic) who would have added to the pollution process and shared their own pathogens.

In short, hunter-gatherers most likely had few pathogenic boarders to purloin a portion of their nutritional intake and few illnesses to fight, with the latter also sapping that intake. Moreover, although no one questions that hunter-gatherers endured hungry

times, their diets in good times featured such a wide variety of nutriments that a healthy mix of nutrients in adequate amounts was ensured.

Sedentism turned this salubrious world upside down. Because their livelihood depended on mobility – on following the food supply – hunter-gatherers produced relatively few children. By contrast, their sedentary successors, who needed hands for the fields and security in old age, reproduced without restraint, and populations began to swell. Squalid villages became even more squalid towns, where people lived cheek to jowl with their growing stock of animals and where diseases began to thrive, along with swarms of insects and rodents that moved in to share in the bounty generated by closely packed humans and their animals.

But even as pathogens were laying an ever-increasing claim to people's nutritional intake, the quality of that intake was sharply declining. The varied diet of hunter-gatherers bore little resemblance to the monotonous diet of their farmer successors, which was most likely to center too closely on a single crop such as wheat, millet, rice, or maize and to feature too little in the way of good-quality protein.

The chapters in Part I focus on this transition, the Neolithic revolutions, which although separated in both time and space, had remarkably similar negative effects on human health.

I.1. 🦋 Dietary Reconstruction and Nutritional Assessment of Past Peoples: The Bioanthropological Record

The topics of diet (the foods that are eaten) and nutrition (the way that these foods are used by the body) are central to an understanding of the evolutionary journey of humankind. Virtually every major anatomical change wrought by that journey can be related in one way or another to how foods are acquired and processed by the human body. Indeed, the very fact that our humanlike ancestors had acquired a bipedal manner of walking by some five to eight million years ago is almost certainly related to how they acquired food. Although the role of diet and nutrition in human evolution has generally come under the purview of anthropology, the subject has also been of great interest to scholars in many other disciplines, including the medical and biological sciences, chemistry, economics, history, sociology, psychology, primatology, paleontology, and numerous applied fields (e.g., public health, food technology, government services). Consideration of nutriture, defined as “the state resulting from the balance between supply of nutrition on the one hand and the expenditure of the organism on the other,” can be traced back to the writings of Hippocrates and Celsus and represents an important heritage of earlier human cultures in both the Old and New Worlds (McLaren 1976, quoted in Himes 1987: 86).

The purpose of this chapter is threefold: (1) to present a brief overview of the basic characteristics of human nutriture and the history of human diet; (2) to examine specific means for reconstructing diet from analysis of human skeletal remains; and (3) to review how the quality of nutrition has been assessed in past populations using evidence garnered by many researchers from paleopathological and skeletal studies and from observations of living human beings. (See also Wing and Brown 1979; Huss-Ashmore, Goodman, and Armelagos 1982; Goodman, Martin, et al. 1984; Martin, Goodman, and Armelagos 1985; Ortner and Putschar 1985; Larsen 1987; Cohen 1989; Stuart-Macadam 1989. For a review of experimental evidence and its implications for humans, see Stewart 1975.) Important developments regarding nutrition in living humans are presented in a number of monographic series, including *World Review of Nutrition and Dietetics*, *Annual Review of Nutrition*, *Nutrition Reviews*, and *Current Topics in Nutrition and Disease*.

Human Nutriture and Dietary History

Although as living organisms we consume foods, we must keep in mind that it is the *nutrients* contained in these foods that are necessary for all of our bodily functions, including support of normal growth and maturation, repair and replacement of body tissues,

and the conduct of physical activities (Malina 1987). Estimations indicate that modern humans require some 40 to 50 nutrients for proper health and well-being (Mann 1981). These nutrients are typically divided into six classes – carbohydrates, proteins, fats, vitamins, minerals, and water. Carbohydrates and fats are the primary energy sources available to the body. Fats are a highly concentrated source of energy and are stored in the body to a far greater degree than carbohydrates. Fats are stored in the range between about 15 and 30 percent of body weight (Malina 1987), whereas carbohydrates represent only about 0.4 to 0.5 percent of body weight in childhood and young adulthood (Fomon et al. 1982). Proteins, too, act as energy sources, but they have two primary functions: tissue growth, maintenance, and repair; and physiological roles.

The building blocks of proteins are chains of nitrogen-containing organic compounds called amino acids. Most of the 22 amino acids can be produced by the body at a rate that is necessary for the synthesis of proteins, and for this reason they are called nonessential amino acids. Eight, however, are not produced in sufficient amounts and therefore must be supplied to the body as food (essential amino acids). Moreover, all essential amino acids have to be present simultaneously in correct amounts and consumed in the same meal in order to be absorbed properly. As noted by W.A. Stini (1971: 1021), “a reliance on any one or combination of foods which lacks even one of the essential amino acids will preclude the utilization of the rest, resulting in continued and increased excretion of nitrogen without compensatory intake.”

Vitamins, a group of 16 compounds, are required in very small amounts only. Save for vitamin D, none of these substances can be synthesized by the body, and if even one is missing or is poorly absorbed, a deficiency disease will arise. Vitamins are mostly regulatory in their overall function. Minerals are inorganic elements that occur in the human body either in large amounts (e.g., calcium and phosphorus) or in trace amounts (called trace elements: e.g., strontium, zinc, fluorine). They serve two important types of functions, namely structural, as in bone and blood production, and regulatory, such as proper balance of electrolytes and fluids. Water, perhaps the most important of the nutrients, functions as a major structural component of the body in temperature regulation and as a transport medium, including elimination of body wastes. About two-thirds of body weight in humans is water (Malina 1987).

Throughout the course of evolution, humans, by adaptation, have acquired a tremendous range of means for securing foods and maintaining proper nutriture. These adaptations can be ordered into a temporal sequence of three phases in the evolution of the human diet (following Gordon 1987). The first phase involved the shift from a diet comprised primarily of unprocessed plant foods to one that incorporated deliberate food-processing techniques and included significant amounts of meat. These changes

likely occurred between the late Miocene epoch and early Pleistocene (or by about 1.5 million years ago). Archaeological and taphonomic evidence indicates that the meat component of diet was likely acquired through a strategy involving scavenging rather than deliberate hunting. Pat Shipman (1986a, 1986b) has examined patterns of cut marks produced by stone tools and tooth marks produced by carnivores in a sample of faunal remains recovered from Olduvai Bed I dating from 2.0 to 1.7 million years ago. In instances where cut marks and tooth marks overlapped on a single bone, her analysis revealed that carnivore tooth marks were *followed* in sequence by hominid-produced cut marks. This pattern of bone modification indicates that hominids scavenged an animal carcass killed by another animal.

The second phase in the history of human diet began in the Middle Pleistocene epoch, perhaps as long ago as 700,000 years before the present. This phase is characterized by deliberate hunting of animal food sources. In East Africa, at the site of Olgorgesailie (700,000 to 400,000 years ago), an extinct species of giant gelada baboon (*Theropithecus oswaldi*) was hunted. Analysis of the remains of these animals by Shipman and co-workers (1981) indicates that although the deaths of many were not due to human activity, young individuals were selectively killed and butchered by hominids for consumption.

Some of the most frequently cited evidence for early hominid food acquisition is from the Torralba and Ambrona sites, located in the province of Soria, Spain (Howell 1966; Freeman 1981). Based on an abundance of remains of large mammals such as elephants, along with stone artifacts, fire, and other evidence of human activity, F. Clark Howell and Leslie G. Freeman concluded that the bone accumulations resulted from "deliberate game drives and the killing of large herbivores by Acheulian hunting peoples" (1982: 13). Richard G. Klein (1987, 1989), however, subsequently argued on the basis of his more detailed observations of animal remains from these sites that despite a human presence as evidenced by stone tools, it is not possible to distinguish between human or carnivore activity in explaining the extensive bone accumulations. First, the relatively greater frequency of axial skeletal elements (e.g., crania, pelves, vertebrae) could be the result of the removal of meatier portions of animal carcasses by either humans or the large carnivores who frequented the site. Second, the overabundance of older elephants could represent human hunting, but it also could represent carnivore activity or natural mortality. Thus, although hominids in Spain were quite likely acquiring protein from animal sources, the evidence based on these Paleolithic sites is equivocal. We know that early hominids acquired meat through hunting activity, but their degree of success in this regard is still unclear.

By later Pleistocene times (20,000 to 11,000 years ago), evidence for specialized hunting strategies clearly

indicates that human populations had developed means by which larger species of animals were successfully hunted. For example, at the Upper Paleolithic site of Solutré, France, Howell (1970) noted that some 100,000 individuals of horse were found at the base of the cliff, and at Predmosti, Czechoslovakia, about 1,000 individuals of mammoth were found. Presumably, the deaths of these animals resulted from purposeful game drives undertaken by local communities of hominids. Virtually all faunal assemblages studied by archaeologists show that large, gregarious herbivores, such as the woolly mammoth, reindeer, bison, and horse, were emphasized, particularly in the middle latitudes of Eurasia (Klein 1989). But some of the best evidence for advances in resource exploitation by humans is from the southern tip of Africa. In this region, Late Stone Age peoples fished extensively, and they hunted dangerous animals like wild pigs and buffalo with considerable success (Klein 1989).

Because of the relatively poor preservation of plant remains as compared to animal remains in Pleistocene sites, our knowledge of the role of plant foods in human Paleolithic nutrition is virtually nonexistent. There is, however, limited evidence from a number of localities. For example, at the *Homo erectus* site of Zhoukoudian in the People's Republic of China (430,000 to 230,000 years before the present), hackberry seeds may have been roasted and consumed. Similarly, in Late Stone Age sites in South Africa, abundant evidence exists for the gathering of plant staples by early modern *Homo sapiens*. Based on what is known about meat and plant consumption by living hunter-gatherers, it is likely that plant foods contributed substantially to the diets of earlier, premodern hominids (Gordon 1987). Today, with the exception of Eskimos, all-meat diets are extremely rare in human populations (Speth 1990), and this almost certainly was the case in antiquity.

The third and final phase in the history of human diet began at the interface between the Pleistocene and Holocene epochs about 10,000 years ago. This period of time is marked by the beginning of essentially modern patterns of climate, vegetation, and fauna. The disappearance of megafauna, such as the mastodon and the mammoth, in many parts of the world at about this time may have been an impetus for human populations to develop new means of food acquisition in order to meet protein and fat requirements. The most important change, however, was the shift from diets based exclusively on food collection to those based to varying degrees on food production.

The transition involved the acquisition by human populations of an intimate knowledge of the life cycles of plants and animals so as to control such cycles and thereby ensure the availability of these nutrients for dietary purposes. By about 7,000 years ago, a transition to a plant-based economy was well established in some areas of the Middle East. From this region, agriculture spread into Europe, and other

independent centers of plant domestication appeared in Africa, Asia, and the New World, all within the next several millennia.

It has been both the popular and scientific consensus that the shift from lifeways based exclusively on hunting and gathering to those that incorporated food production – and especially agriculture – represented a positive change for humankind. However, Mark N. Cohen (1989) has remarked that in game-rich environments, regardless of the strategy employed, hunters may obtain between 10,000 and 15,000 kilocalories per hour. Subsistence cultivators, in contrast, average between 3,000 and 5,000 kilocalories per hour.

More important, anthropologists have come to recognize in recent years that the shift from hunting and gathering to agriculture was characterized by a shift from generally high-quality foods to low-quality foods. For example, animal sources of protein contain all essential amino acids in the correct proportions. They are a primary source of vitamin B₁₂, are high in vitamins A and D, and contain important minerals. Moreover, animal fat is a critical source of essential fatty acids and fat-soluble vitamins (Speth 1990). Thus, relative to plant foods, meat is a highly nutritional food resource. Plant foods used alone generally cannot sustain human life, primarily because of deficiency in essential amino acids (see discussion in Ross 1976). Moreover, in circumstances where plant foods are emphasized, a wide variety of them must be consumed in order to fulfill basic nutritional requirements. Further limiting the nutritional value of many plants is their high fiber content, especially cellulose, which is not digestible by humans.

Periodic food shortages resulting from variation in a number of factors – especially rainfall and temperature, along with the relative prevalence of insects and other pests – have been observed in contemporary human populations depending on subsistence agriculture. Some of the effects of such shortages include weight loss in adults, slowing of growth in children, and an increase in prevalence of malaria and other diseases such as gastroenteritis, and parasitic infection (Bogin 1988).

Archaeological evidence from prehistoric agriculturalists, along with observation of living peasant agriculturalists, indicates that their diets tend to be dominated by a single cereal staple: rice in Asia, wheat in temperate Asia and Europe, millet or sorghum in Africa, and maize in the New World. These foods are oftentimes referred to as *superfoods*, not because of nutritional value but rather because of the pervasive focus by human populations on one or another of them (McElroy and Townsend 1979).

Rice, a food staple domesticated in Southeast Asia and eventually extending in use from Japan and Korea southward to Indonesia and eastward into parts of India, has formed the basis of numerous complex cultures and civilizations (Bray 1989). Yet it is remarkably deficient in protein, even in its brown or unmilled

form. Moreover, the low availability of protein in rice inhibits the activity of vitamin A, even if the vitamin is available through other food sources (Wolf 1980). Vitamin A deficiency can trigger xerophthalmia, one of the principal causes of blindness. White rice – the form preferred by most human populations – results from processing, or the removal of the outer bran coat, and consequently, the removal of thiamine (vitamin B₁). This deficiency leads to beriberi, a disease alternately involving inflammation of the nerves, or the heart, or both.

Wheat was domesticated in the Middle East very early in the Holocene and has been widely used since that time. Wheat is deficient in two essential amino acids – lysine and isoleucine. Most human populations dependent on wheat, however, have dairy animals that provide products (e.g., cheese) that make up for these missing amino acids. Yet in some areas of the Middle East and North Africa where wheat is grown, zinc-deficient soils have been implicated in retarding growth in children (Harrison et al. 1988). Moreover, the phytic acid present in wheat bran chemically binds with zinc, thus inhibiting its absorption (Mottram 1979).

Maize (known as corn in the United States) was first domesticated in Mesoamerica. Like the other superfoods, it formed the economic basis for the rise of civilizations and complex societies, and its continued domestication greatly increased its productivity (Galinat 1985). In eastern North America, maize was central in the evolution of a diversity of chiefdoms (Smith 1989), and its importance in the Americas was underscored by Walton C. Galinat, who noted:

[By] the time of Columbus, maize had already become the staff of life in the New World. It was distributed throughout both hemispheres from Argentina and Chile northward to Canada and from sea level to high in the Andes, from swampland to arid conditions and from short to long day lengths. In becoming so widespread, it evolved hundreds of races, each with special adaptations for the environment including special utilities for man. (Galinat 1985: 245)

Like the other superfoods, maize is deficient in a number of important nutrients. Zein – the protein in maize – is deficient in lysine, isoleucine, and tryptophan (FAO 1970), and if maize consumers do not supplement their diets with foods containing these amino acids, such as beans, significant growth retardation is an outcome. Moreover, maize, although not deficient in niacin (vitamin B₃), contains it in a chemically bound form that, untreated, will withhold the vitamin from the consumer. Consequently, human populations consuming untreated maize frequently develop pellagra, a deficiency disease characterized by a number of symptoms, including rough and irritated skin, mental symptoms, and diarrhea (Roe 1973). Solomon H. Katz and co-workers (1974, 1975;

see also Katz 1987) have shown that many Native American groups treat maize with alkali (e.g., lye, lime, or wood ashes) prior to consumption, thereby liberating niacin. Moreover, the amino acid quality in alkali-treated maize is significantly improved. Most human populations who later acquired maize as a dietary staple did not, however, adopt the alkali-treatment method (see Roe 1973). Maize also contains phytate and sucrose, whose negative impact on human health is considered later in this chapter.

Dietary Reconstruction: Human Remains

Chemistry and Isotopy

Skeletal remains from archaeological sites play a very special role in dietary reconstruction because they provide the only *direct* evidence of food consumption practices in past societies. In the last decade, several trace elements and stable isotopes have been measured and analyzed in human remains for the reconstruction of diets. Stanley H. Ambrose (1987) has reviewed these approaches, and the following is drawn from his discussion (see also van der Merwe 1982; Klepinger 1984; Sealy 1986; Aufderheide 1989; Keegan 1989; Schoeninger 1989; Sandford 1993).

Some elements have been identified as potentially useful in dietary reconstruction. These include manganese (Mn), strontium (Sr), and barium (Br), which are concentrated in plant foods, and zinc (Zn) and copper (Cu), which are concentrated in animal foods. Nuts, which are low in vanadium (V), contrast with other plant foods in that they typically contain high amounts of Cu and Zn. Like plants, marine resources (e.g., shellfish) are usually enriched in Sr, and thus the dietary signatures resulting from consumption of plants and marine foods or freshwater shellfish should be similar (Schoeninger and Peebles 1981; Price 1985). In contrast, Br is deficient in bones of marine animals, thereby distinguishing these organisms from terrestrial ones in this chemical signature (Burton and Price 1990).

The greater body of research done on elemental composition has been with Sr. In general, Sr levels decline as one moves up the food chain – from plants to herbivores to primary carnivores – as a result of natural biopurification (a process called fractionation). Simply put, herbivores consume plants that are enriched with Sr contained in soil. Because very little of the Sr that passes through the gut wall in animals is stored in flesh (only about 10 percent), the carnivore consuming the herbivore will have considerably less strontium stored in its skeleton. Humans and other omnivores, therefore, should have Sr concentrations that are intermediate between herbivores and carnivores in their skeletal tissues. Thus, based on the amount of Sr measured in human bones, it is possible (with some qualifications) to determine the relative contributions of plant and meat foods to a diet.

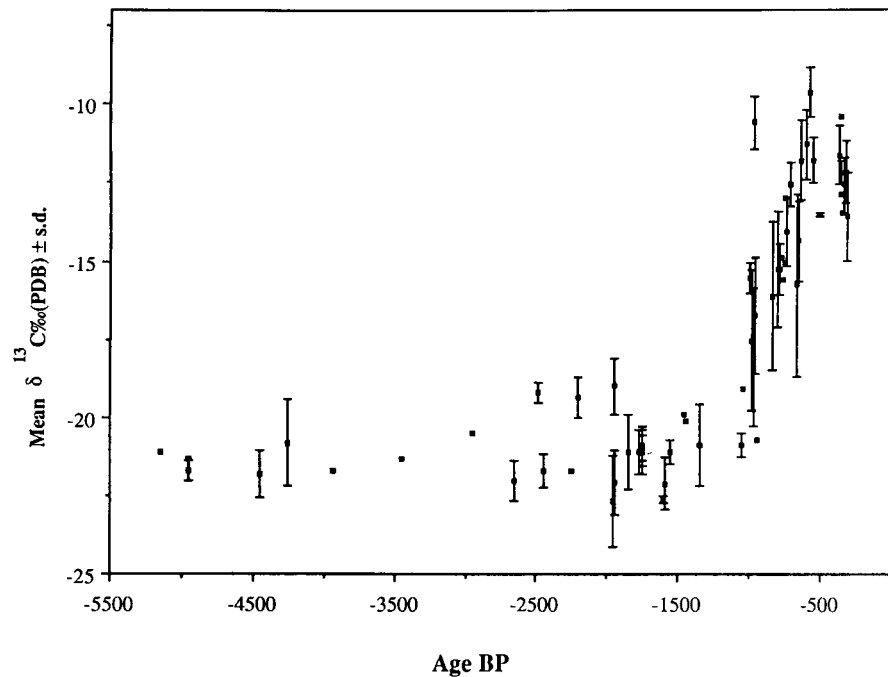
Nonetheless, in addition to the aforementioned

problem with shellfish, there are three chief limitations to Sr and other elemental analyses. First, Sr abundance can vary widely from region to region, depending upon the geological context. Therefore, it is critical that the baseline elemental concentrations in local soils – and plants and animals – be known. Second, it must be shown that diagenesis (the process involving alteration of elemental abundance in bone tissue while it is contained in the burial matrix) has not occurred. Some elements appear to resist diagenesis following burial (e.g., Sr, Zn, Pb [lead], Na [sodium]), and other elements show evidence for diagenesis (e.g., Fe [iron], Al [aluminum], K [potassium], Mn, Cu, Ba). Moreover, diagenetic change has been found to vary within even a single bone (e.g., Sillen and Kavanagh 1982; Bumsted 1985). Margaret J. Schoeninger and co-workers (1989) have evaluated the extent of preservation of histological structures in archaeological bone from the seventeenth-century Georgia coastal Spanish mission Santa Catalina de Guale. This study revealed that bones with the least degree of preservation of structures have the lowest Sr concentrations. Although these low values may result from diet, more likely they result from diagenetic effects following burial in the soil matrix. And finally, pretreatment procedures of archaeological bone samples in the laboratory frequently are ineffective in completely removing the contaminants originating in groundwater, such as calcium carbonate, thus potentially masking important dietary signatures.

Valuable information on specific aspects of dietary composition in past human populations can also be obtained by the analysis of stable isotopes of organic material (collagen) in bone. Isotopes of two elements have proven of value in the analysis of diets: carbon (C) and nitrogen (N). Field and laboratory studies involving controlled feeding experiments have shown that stable isotope ratios of both carbon ($^{13}\text{C}/^{12}\text{C}$) and nitrogen ($^{15}\text{N}/^{14}\text{N}$) in an animal's tissues, including bone, reflect similar ratios of diet. Because the variations in isotopic abundances between dietary resources are quite small, the values in tissue samples are expressed in parts per thousand (o/oo) relative to established standards, as per delta (δ) values.

The $\delta^{13}\text{C}$ values have been used to identify two major dietary categories. The first category has been used to distinguish consumers of plants with different photosynthetic pathways, including consumers of C4 plants (tropical grasses such as maize) and consumers of C3 plants (most leafy plants). Because these plants differ in their photosynthetic pathways, they also differ in the amount of ^{13}C that they incorporate. Thus, C4 plants and people who consume them have $\delta^{13}\text{C}$ values that differ on average by about 14 o/oo from other diets utilizing non-C4 plants. Based on these differences, it has been possible to track the introduction and intensification of maize agriculture in eastern North America with some degree of precision (Figure I.1.1). The second cate-

Figure I.1.1. Temporal changes in mean values of $\delta^{13}\text{C}$ of prehistoric eastern North American Indians. The error bars represent one standard deviation. (From Ambrose 1987, in *Emergent Horticultural Economies of the Eastern Woodlands*, ed. W. F. Keegan; ©1987 the Board of Trustees, Southern Illinois University; reprinted by permission of the author and the Center for Archaeological Investigations.)



gory of dietary identification reflected by $\delta^{13}\text{C}$ values includes primarily marine foods. Marine fish and mammals have more positive $\delta^{13}\text{C}$ values (by about 6 ‰) compared to terrestrial animals feeding on C3 foods, and less positive values (by about 7 ‰) than terrestrial animals feeding on C4 foods (especially maize) (Schoeninger and DeNiro 1984; Schoeninger, van der Merwe, and Moore 1990).

Nitrogen stable isotope ratios in human bone are used to distinguish between consumers of terrestrial and marine foods. Margaret Schoeninger and Michael J. DeNiro (1984; see also Schoeninger et al. 1990) have indicated that in many geographical regions the $\delta^{15}\text{N}$ values of marine organisms differ from terrestrial organisms by about 10 parts per thousand on average, with consumers of terrestrial foods being less positive than consumers of marine foods. Recent research on stable isotopes of sulphur (i.e., $^{34}\text{S}/^{32}\text{S}$) suggests that they may provide an additional means of identifying diets based on marine foods from those based on terrestrial foods because of the relatively greater abundance of ^{34}S in marine organisms (Krouse 1987). A preliminary study of prehistoric populations from coastal Chile has supported this distinction in human remains representative of marine and terrestrial subsistence economies (Kelley, Levesque, and Weidl 1991).

As already indicated, based on carbon stable isotope values alone, the contribution of maize to diets in populations consuming marine resources is difficult to assess from coastal areas of the New World because of the similarity of isotope signatures of marine foods

and individuals with partial maize diets (Schoeninger et al. 1990). However, by using both carbon and nitrogen isotope ratios, it is possible to distinguish between the relative contributions to diets of marine and terrestrial (e.g., maize) foods (Schoeninger et al. 1990).

Stable isotopes (C and N) have several advantages over trace elements in dietary documentation. For example, because bone collagen is not subject to isotopic exchange, diagenetic effects are not as important a confounding factor as in trace elemental analysis (Ambrose 1987; Grupe, Piepenbrink, and Schoeninger 1989). Perhaps the greatest advantage, however, is that because of the relative ease of removing the mineral component of bone (as well as fats and humic contaminants) and of confirming the collagen presence through identification of amino acids, the sample purity can be controlled (Ambrose 1987; Stafford, Brendel, and Duhamel 1988). However, collagen abundance declines in the burial matrix, and it is the first substance to degrade in bone decomposition (Grupe et al. 1989). If the decline in collagen value does not exceed 5 percent of the original value, then the isotopic information is suspect (see also Bada, Schoeninger, and Schimmelmann 1989; Schoeninger et al. 1990). Therefore, human fossil specimens, which typically contain little or no collagen, are generally not conducive to dietary reconstruction.

Teeth and Diet: Tooth Wear

Humankind has developed many means of processing foods before they are eaten. Nevertheless, virtually all foods have to be masticated by use of the teeth to one

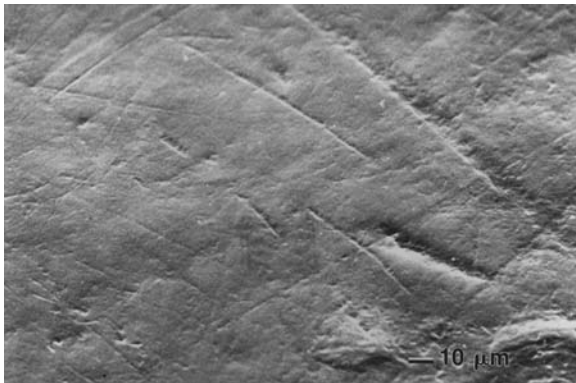
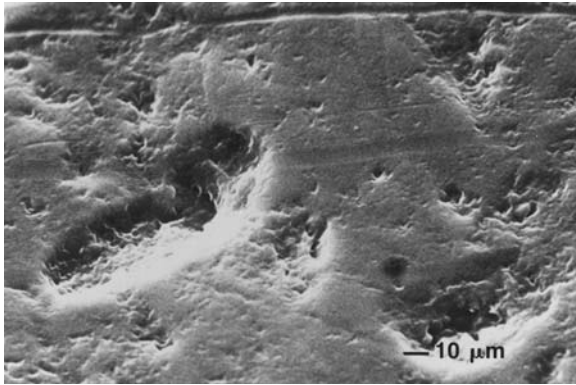


Figure I.1.2. Scanning electron micrographs ($\times 500$) of prehistoric hunter-gatherer molar (top) and historic agriculturalist molar (bottom) from the southeastern U.S. Atlantic coast. (From Teaford 1991, in *Advances in Dental Anthropology*, ed. Marc A. Kelley and Clark Spencer Larsen; ©1991; reprinted by permission of the author and Wiley-Liss, a division of John Wiley and Sons, Inc.)

extent or another before they are passed along for other digestive activities. Because food comes into contact with teeth, the chewing surfaces of teeth wear. Defined as “the loss of calcified tissues of a tooth by erosion, abrasion, attrition, or any combination of these” (Wallace 1974: 385), tooth wear – both microscopic and macroscopic – provides information on diets of past populations. The importance of tooth wear in the reconstruction of diet has been underscored by Phillip L. Walker (1978: 101), who stated, “From an archaeological standpoint, dietary information based on the analysis of dental attrition is of considerable value since it offers an independent check against reconstruction of prehistoric subsistence based on the analysis of floral, faunal and artifactual evidence.”

Recent work with use of scanning electron microscopy (SEM) in the study of microwear on occlusal surfaces of teeth has begun to produce important data on diet in human populations (reviewed in Teaford 1991) (Figure I.1.2). Field and laboratory studies have shown

that microwear features can change rapidly. Therefore, microwear patterns may give information only on food items consumed shortly before death. These features, nevertheless, have been shown to possess remarkable consistency across human populations and various animal species and have, therefore, provided insight into past culinary habits. For example, hard-object feeders, including Miocene apes (e.g., *Sivapithecus*) as well as recent humans, consistently develop large pits on the chewing surfaces of teeth. In contrast, consumers of soft foods, such as certain agriculturalists (Bullington 1988; Teaford 1991), develop smaller and fewer pits as well as narrower and more frequently occurring scratches.

Macroscopic wear can also vary widely, depending upon a host of factors (Molnar 1972; Foley and Cruwys 1986; Hillson 1986; Larsen 1987; Benfer and Edwards 1991; Hartnady and Rose 1991; Walker, Dean, and Shapiro 1991). High on the list of factors affecting wear, however, are the types of foods consumed and manner of their preparation. Because most Western populations consume soft, processed foods with virtually all extraneous grit removed, tooth wear occurs very slowly. But non-Western populations consuming traditional foods (that frequently contain grit contaminants introduced via grinding stones) show rapid rates of dental wear (e.g., Hartnady and Rose 1991). Where there are shifts in food types (e.g., from hunting and gathering to agriculture) involving reduction in food hardness or changes in how these foods are processed (e.g., with stone versus wooden grinding implements), most investigators have found a reduction in gross wear (e.g., Anderson 1965, 1967; Walker 1978; Hinton, Smith, and Smith 1980; Smith, Smith, and Hinton 1980; Patterson 1984; Bennike 1985; Inoue, Ito, and Kamegai 1986; Benfer and Edwards 1991; Rose, Marks, and Tieszen 1991).

Consistent with reductions in tooth wear in the shift to softer diets are reductions in craniofacial robusticity, both in Old World settings (e.g., Carlson and Van Gerven 1977, 1979; Armelagos, Carlson, and Van Gerven 1982; y’Edynak and Fleisch 1983; Smith, Bar-Yosef, and Sillen 1984; Wu and Zhang 1985; Inoue et al. 1986; y’Edynak 1989) and in New World settings (e.g., Anderson 1967; Larsen 1982; Boyd 1988). In prehistoric Tennessee Amerindians, for example, Donna C. Boyd (1988) has documented a clear trend for a reduction in dimensions of the mandible and facial bones that reflects decreasing masticatory stress relating to a shift to soft foods. Although not all studies of this sort examine both craniofacial and dental wear changes, those that do so report reductions in both craniofacial robusticity and dental wear, reflecting a decrease in hardness of foods consumed (Anderson 1967; Inoue et al. 1986). Other changes accompanying shifts from hard-textured to soft-textured foods include an increase in malocclusion and crowding of teeth due to inadequate growth of the jaws (reviewed by Corruccini 1991).

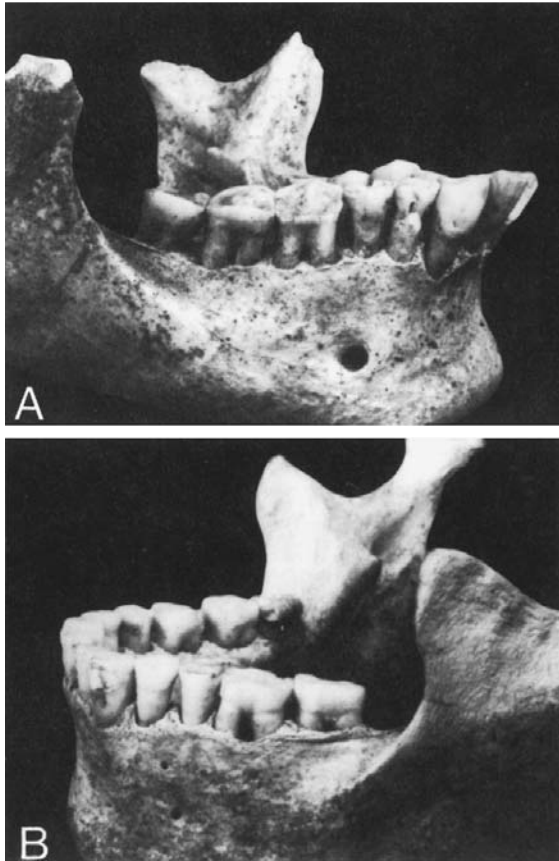


Figure I.1.3. Views of mandibular dentitions showing agriculturalist (A) and hunter-gatherer (B) wear planes. Note the angled wear on the agriculturalist's molars and the flat wear on the hunter-gatherer's molars. (From Smith 1984, in *American Journal of Physical Anthropology*; ©1984; reprinted by permission of the author and Wiley-Liss, a division of John Wiley and Sons, Inc.)

B. Holly Smith (1984, 1985) has found consistent patterns of tooth wear in human populations (Figure I.1.3). In particular, agriculturalists – regardless of regional differences – show highly angled molar wear planes in comparison with those of hunter-gatherers. The latter tend to exhibit more evenly distributed, flat wear. Smith interpreted the differences in tooth wear between agriculturalists and hunter-gatherers as reflecting greater “toughness” of hunter-gatherer foods.

Similarly, Robert J. Hinton (1981) has found in a large series of Native American dentitions representative of hunter-gatherers and agriculturalists that the former wear their anterior teeth (incisors and canines) at a greater rate than the latter. Agriculturalists that he studied show a tendency for cupped wear on the chewing surfaces of the anterior teeth. Because agriculturalists exhibit a relatively greater rate of premortem posterior tooth loss (especially molars), Hinton relates

the peculiar wear pattern of the anterior teeth to the use of these teeth in grinding food once the molars are no longer available for this masticatory activity.

Specific macroscopic wear patterns appear to arise as a result of chewing one type of food. In a prehistoric population from coastal Brazil, Christy G. Turner II and Lilia M. Machado (1983) found that in the anterior dentition the tooth surfaces facing the tongue were more heavily worn than the tooth surfaces facing the lips (Figure I.1.4.). They interpreted this wear pattern as reflecting the use of these teeth to peel or shred abrasive plants for dietary or extra-masticatory purposes.

Teeth and Diet: Dental Caries

The health of the dental hard tissues and their supporting bony structures are intimately tied to diet. Perhaps the most frequently cited disease that has been linked with diet is dental caries, which is defined as “a disease process characterized by the focal demineralization of dental hard tissues by organic acids produced by bacterial fermentation of dietary carbohydrates, especially sugars” (Larsen 1987: 375). If the decay of tooth crowns is left unchecked, it will lead to cavitation, loss of the tooth, and occasionally, infection and even death (cf. Calcagno and Gibson 1991). Carious lesions can develop on virtually any exposed surface of the tooth crown. However, teeth possessing grooves and fissures (especially posterior teeth) tend to trap food particles and are, therefore, more prone to colonization by indigenous bacteria, and thus to cariogenesis. Moreover, pits and linear depressions arising from poorly formed enamel (hypoplasia or hypocalcification) are also predisposed to caries attack, especially in populations with cariogenic diets (Powell 1985; Cook 1990) (Figure I.1.5).



Figure I.1.4. Lingual wear on anterior teeth of prehistoric Brazilian Indian. (From Turner and Machado 1983; in *American Journal of Physical Anthropology*; ©1984; reprinted by permission of the authors and Wiley-Liss, a division of John Wiley and Sons, Inc.)

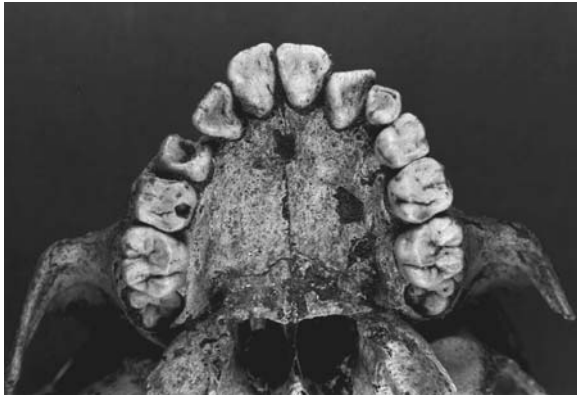


Figure I.1.5. Dental carious lesion in maxillary molar from historic Florida Indian. (Photograph by Mark C. Griffin.)

Dental caries is a disease with considerable antiquity in humans. F. E. Grine, A. J. Gwynnett, and J. H. Oaks (1990) note the occurrence of caries in dental remains of early hominids dating from about 1.5 million years ago (robust australopithecines and *Homo erectus*) from the Swartkrans site (South Africa), albeit at low prevalence levels. Later *Homo erectus* teeth from this site show higher prevalence than australopithecines, which may reflect their consumption of honey, a caries-promoting food (Grine et al. 1990). But with few exceptions (e.g., the Kabwe early archaic *Homo sapiens* from about 130,000 years before the present [Brothwell 1963]), caries prevalence has been found to be very low until the appearance of plant domestication in the early Holocene. David W. Frayer (1988) has documented one of these exceptions – an unusually high prevalence in a Mesolithic population from Portugal, which he relates to the possible consumption of honey and figs.

Turner (1979) has completed a worldwide survey of archaeological and living human populations whereby diet has been documented and the percentage of carious teeth has been tabulated. The samples were subdivided into three subsistence groups: hunting and gathering ($n = 19$ populations), mixed (combination of agriculture with hunting, gathering, or fishing; $n = 13$ populations), and agriculture ($n = 32$ populations). By pooling the populations within each subsistence group, Turner found that hunter-gatherers exhibited 1.7 percent carious teeth, mixed subsistence groups (combining hunting, gathering, and agriculture) exhibited 4.4 percent carious teeth, and agriculturalists exhibited 8.6 percent carious teeth.

Other researchers summarizing large comparative samples have confirmed these findings, especially with regard to a dichotomy in caries prevalence between hunter-gatherers and agriculturalists. Clark Spencer Larsen and co-workers (1991) compared 75 archaeological dental samples from the eastern United States. Only three agriculturalist populations

exhibited less than 7 percent carious teeth, and similarly, only three hunter-gatherer populations exhibited greater than 7 percent carious teeth. The greater frequencies of carious teeth in the agricultural populations are largely due to those people's consumption of maize (see also Milner 1984). The cariogenic component of maize is sucrose, a simple sugar that is more readily metabolized by oral bacteria than are more complex carbohydrates (Newbrun 1982). Another factor contributing to high caries prevalence in later agricultural populations may be due to the fact that maize is frequently consumed in the form of soft mushes. These foods have the tendency to become trapped in grooves and fissures of teeth, thereby enhancing the growth of plaque and contributing to tooth decay due to the metabolism of sugar by indigenous bacteria (see also Powell 1985).

High prevalence of dental caries does not necessarily indicate a subsistence regime that included maize agriculture, because other carbohydrates have been strongly implicated in prehistoric nonagricultural contexts. Philip Hartnady and Jerome C. Rose (1991) reported a high frequency of carious lesions – 14 percent – in the Lower Pecos region of southwest Texas. These investigators related elevated levels of caries to the consumption of plants high in carbohydrates, namely sotol, prickly pear, and lecheguilla. The fruit of prickly pear (known locally as tuna) contains a significant sucrose component in a sticky, pectin-based mucilage. The presence of a simple sugar in this plant food, coupled with its gummy nature, is clearly a caries-promoting factor (see also Walker and Erlandson 1986, and Kelley et al. 1991, for different geographical settings involving consumption of nonagricultural plant carbohydrates).

Nutritional Assessment

Growth and Development

One of the most striking characteristics of human physical growth during the period of infancy and childhood is its predictability (Johnston 1986; Bogin 1988). Because of this predictability, anthropometric approaches are one of the most commonly used indices in the assessment of health and well-being, including nutritional status (Yarbrough et al. 1974). In this regard, a number of growth standards based on living subjects have been established (Gracey 1987). Comparisons of individuals of known age with these standards make it possible to identify deviations from the “normal” growth trajectory.

Growth is highly sensitive to nutritional quality, especially during the earlier years of infancy and early childhood (birth to 2 years of age) when the human body undergoes very rapid growth. The relationship between nutrition and growth has been amply demonstrated by the observation of recent human populations experiencing malnutrition. These populations show a secular trend for reduced physical size

of children and adults followed by increased physical size with improvements in diet (e.g., for Japanese, see Kimura 1984; Yagi, Takebe, and Itoh 1989; and for additional populations, Eveleth and Tanner 1976).

Based on a large sample of North Americans representative of different socioeconomic groups, Stanley M. Garn and co-workers (Garn, Owen, and Clark 1974; Garn and Clark 1975) reported that children in lower income families were shorter than those in higher income families (see also review in Bogin 1988). Although a variety of factors may be involved, presumably the most important is nutritional status.

One means of assessing nutrition and its influence on growth and development in past populations is by the construction of growth curves based on comparison of length of long bones in different juvenile age groups (e.g., Merchant and Ubelaker 1977; Sundick 1978; Hummert and Van Gerven 1983; Goodman, Lallo, et al. 1984; Jantz and Owsley 1984; Owsley and Jantz 1985; Lovejoy, Russell, and Harrison 1990) (Figure I.1.6). These data provide a reasonable profile of rate or velocity of growth. Della C. Cook (1984), for example, studied the remains of a group ranging in age from birth to 6 years. They were from a time-successive population in the midwestern United States undergoing the intensification of food production and increased reliance on maize agriculture. Her analysis revealed that individuals living during the introduction of maize had shorter femurs for their age than did individuals living before, as hunter-gatherers, or those living after, as maize-intensive agriculturalists (Cook 1984). Analysis of depressed growth among prehistoric hunter-gatherers at the Libben site (Ohio) suggests, however, that infectious disease was a more likely culprit in this context because the hunter-gatherers' nutrition - based on archaeological reconstruction of their diet - was adequate (Lovejoy et al. 1990).

Comparison of skeletal development, a factor responsive to nutritional insult, with dental development, a factor that is relatively less responsive to nutritional insult (see the section "Dental Development"), can provide corroborative information on nutritional status in human populations. In two series of archaeological populations from Nubia, K. P. Moore, S. Thorp, and D. P. Van Gerven (1986) compared skeletal age and dental age and found that most individuals (70.5 percent) had a skeletal age younger than their dental age. These findings were interpreted as reflecting significant retardation of skeletal growth that was probably related to high levels of nutritional stress. Indeed, undernutrition was confirmed by the presence of other indicators of nutritional insult, such as iron-deficiency anemia.

In living populations experiencing generally poor nutrition and health, if environmental insults are removed (e.g., if nutrition is improved), then children may increase in size, thereby more closely approximating their genetic growth potential (Bogin 1988). However, if disadvantageous conditions are sustained,

then it is unlikely that the growth potential will be realized. Thus, despite prolonged growth in undernourished populations, adult height is reduced by about 10 percent (Frisancho 1979). Sustained growth depression occurring during the years of growth and development, then, has almost certain negative consequences for final adult stature (Bogin 1988 and references cited therein).

In archaeological populations, reductions in stature have been reported in contexts with evidence for reduced nutritional quality. On the prehistoric Georgia coast, for example, there was a stature reduction of about 4 centimeters in females and 2 centimeters in males during the shift from hunting, gathering, and fishing to a mixed economy involving maize agriculture (Larsen 1982; Angel 1984; Kennedy 1984; Meiklejohn et al. 1984; Rose et al. 1984; and discussions in Cohen and Armelagos 1984; Larsen 1987; Cohen 1989). All workers documenting reductions in stature regard it as reflecting a shift to the relatively poor diets that are oftentimes associated with agricultural food production such as maize in North America.

Cortical Bone Thickness

Bone tissue, like any other tissue of the body, is subject to environmental influences, including nutritional quality. In the early 1960s, Garn and co-workers (1964) showed that undernourished Guatemalan children had reduced thickness of cortical (sometimes called compact) bone compared with better nourished children from the same region. Such changes were related to loss of bone during periods of acute protein energy malnutrition. These findings have been confirmed by a large number of clinical investigations (e.g., Himes et al. 1975; and discussion in Frisancho 1978).

Bone maintenance in archaeological skeletal populations has been studied by a number of investigators.

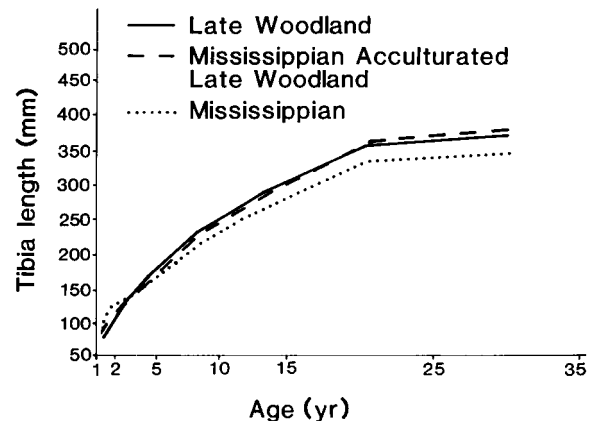


Figure I.1.6. Growth curves from Dickson Mounds, Illinois, Indian population. (Adapted from Lallo 1973; reproduced from Larsen 1987 with permission of Academic Press, Inc.)

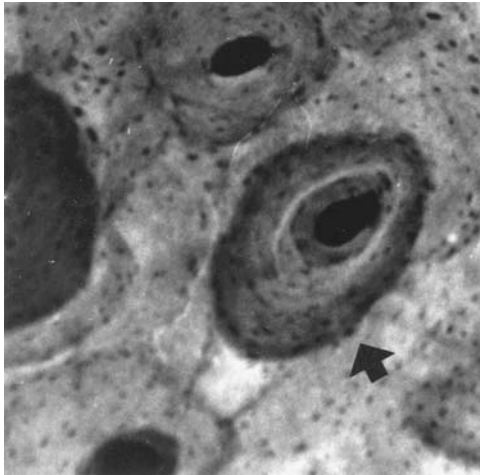


Figure I.1.7. Micrograph ($\times 150$) showing hypermineralized rings (dark zones) within an osteon from prehistoric Nubian. (Photograph courtesy of Debra L. Martin.)

Most frequently expressed as a ratio of the amount of cortical bone to subperiosteal area – or percent cortical area (PCCA) or percent cortical thickness (PCCT) – it has been interpreted by most people working with archaeological human remains as reflecting nutritional or health status (e.g., Cassidy 1984; Cook 1984; Brown 1988; Cohen 1989). It is important to note, however, that bone also remodels itself under conditions of mechanical demand, so that bone morphology that might be interpreted as reflecting a reduction in nutritional status may in fact represent an increase in mechanical loading (Ruff and Larsen 1990).

Cortical Bone Remodeling and Microstructure

An important characteristic that bone shares with other body tissues is that it must renew itself. The renewal of bone tissue, however, is unique in that the process involves destruction followed by replacement with new tissue. The characteristic destruction (resorption) and replacement (deposition) occurs mostly during the years of growth and development prior to adulthood, but it continues throughout the years following. Microstructures observable in bone cross sections have been analyzed and have provided important information about bone remodeling and its relationship to nutritional status. These microstructures include osteons (tunnels created by resorption and partially filled in by deposition of bone tissue), Haversian canals (central canals associated with osteons), and surrounding bone.

As with cortical thickness, there is a loss of bone mass that can be observed via measurement of the degree of porosity through either invasive (e.g., histological bone thin sections) or noninvasive (e.g., photon absorptiometry) means. With advancing age, cortical bone becomes both thinner and more porous. Cortical bone that has undergone a reduction in bone

mass per volume – a disorder called osteoporosis – should reflect the nutritional history of an individual, especially if age factors have been ruled out (Martin et al. 1985; Schaafsma et al. 1987; Arnaud and Sanchez 1990). If this is the case, then bone loss can affect any individual regardless of age (Stini 1990). Clinical studies have shown that individuals with low calcium intakes are more prone to bone loss in adulthood (Nordin 1973; Arnaud and Sanchez 1990; Stini 1990). It is important to emphasize, however, that osteoporosis is a complex, multifactorial disorder and is influenced by a number of risk factors, including nondietary ones such as body weight, degree of physical exercise, and heredity (Evers, Orchard, and Haddad 1985; Schaafsma et al. 1987; Arnaud and Sanchez 1990; Stini 1990; Ruff 1991; Lindsay and Cosman 1992; Heaney 1993).

Porosity of bone also represents a function of both the number of Haversian canals and their size (Atkinson 1964; Thompson 1980; Burr, Ruff, and Thompson 1990). Therefore, the greater the number and width of Haversian canals, the greater the porosity of bone tissue. The density of individual osteons appears to be related to nutritional quality as well. For example, the presence of osteons containing hypermineralized lines in archaeological human remains likely reflects periods of growth disturbance (e.g., Stout and Simmons 1979; Martin and Armelagos 1985) (Figure I.1.7).

Samuel D. Stout and co-workers (Stout and Teitelbaum 1976; Stout 1978, 1983, 1989) have made comparisons of bone remodeling dynamics between a series of hunter-gatherer and maize-dependent North and South American archaeological populations. Their findings show that the single agricultural population used in the study (Ledders, Illinois) had bone remodeling rates that were higher than the other (non-maize) populations. They suggested that because maize is low in calcium and high in phosphorus, parathyroid hormone levels could be increased. Bone remodeling is highly stimulated by parathyroid hormone, a disorder known as hyperparathyroidism.

In order to compensate for bone loss in aging adults (particularly after 40 years of age), there are structural adaptations involving more outward distribution of bone tissue in the limb bones. In older adults, such adaptation contributes to maintaining the biomechanical strength despite bone losses (Ruff and Hayes 1982). Similarly, D. B. Burr and R. B. Martin (1983; see also Burr et al. 1990) have suggested that the previously discussed material property changes may supplement structural changes. Thus, different rates of bone turnover in human populations may reflect mechanical adaptations that are not necessarily linked to poor nutrition.

Skeletal (Harris) Lines of Increased Density

Nonspecific markers of physiological stress that appear to have some links with nutrition status are radiographically visible lines of increased bone den-

sity, referred to as Harris lines (Figure I.1.8). These lines either partly or completely span the medullary cavities of tubular bones (especially long bones of the arms and legs) and trace the outlines of other bones (Garn et al. 1968; Steinbock 1976). These lines have been found to be associated with malnutrition in modern humans (e.g., Jones and Dean 1959) and in experimental animals (e.g., Stewart and Platt 1958). Because Harris lines develop during bone formation, age estimates can be made for time of occurrence relative to the primary ossification centers (e.g., Goodman and Clark 1981). However, the usefulness of these lines for nutritional assessment is severely limited by the fact that they frequently resorb in adulthood (Garn et al. 1968). Moreover, Harris lines have been documented in cases where an individual has not undergone episodes of nutritional or other stress (Webb 1989), and they appear to correlate negatively with other stress markers (reviewed in Larsen 1987).

Dental Development: Formation and Eruption

Like skeletal tissues, dental tissues are highly sensitive to nutritional perturbations that occur during the years of growth and development. Unlike skeletal tissues, however, teeth – crowns, in particular – do not remodel once formed, and they thereby provide a permanent “memory” of nutritional and health history. Alan H. Goodman and Jerome C. Rose (1991: 279) have underscored the importance of teeth in the anthropological study of nutrition: “Because of the inherent and close relationship between teeth and diet, the dental structures have incorporated a variety of characteristics that reflect what was placed in the mouth and presumably consumed” (see also Scott and Turner 1988).

There are two main factors involved in dental development – *formation* of crowns and roots and *eruption* of teeth. Because formation is more heritable than eruption, it is relatively more resistant to nutritional insult (Smith 1991). Moreover, the resistance of formation to environmental problems arising during the growth years is suggested by low correlations between formation and stature, fatness, body weight, or bone age, and lack of secular trend. Thus, timing of formation of tooth crowns represents a poor indicator for assessing nutritional quality in either living or archaeological populations. Eruption, however, can be affected by a number of factors, including caries, tooth loss, and severe malnutrition (e.g., Alvarez et al. 1988, 1990; Alvarez and Navia 1989). In a large, cross-sectional evaluation of Peruvian children raised in nutritionally deprived settings, J. O. Alvarez and co-workers (1988) found that exfoliation of the deciduous dentition was delayed. Other workers have found that eruption was delayed in populations experiencing nutritional deprivation (e.g., Barrett and Brown 1966; Alvarez et al. 1990). Unlike formation, eruption timing has been shown to be correlated with various measures of body size (Garn, Lewis, and Polacheck 1960; McGregor, Thomson, and Billewicz 1968). To my

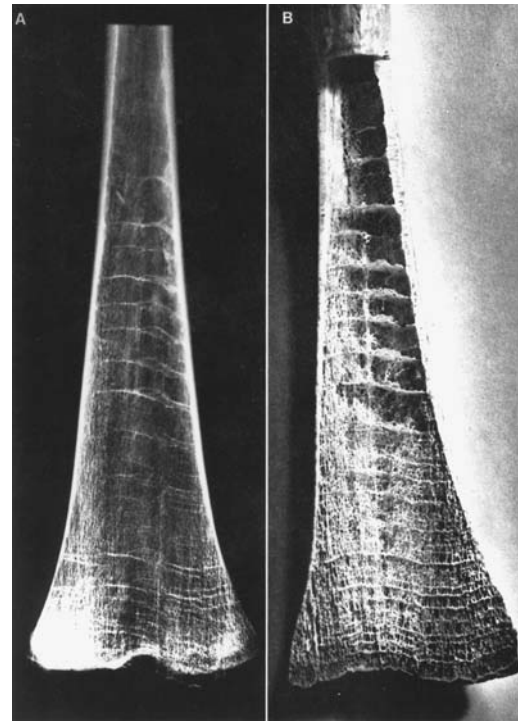


Figure I.1.8. Radiograph (A) and section (B) of prehistoric California Indian femur with Harris lines. (From McHenry 1968, in *American Journal of Physical Anthropology*; ©1968; reprinted by permission of author and Wiley-Liss, a division of John Wiley and Sons, Inc.)

knowledge, there have been no archaeological populations where delayed eruption timing has been related to nutritional status.

Dental Development: Tooth Size

Unlike formation timing, tooth size appears to be under the influence of nutritional status. Garn and co-workers (Garn and Burdi 1971; Garn, Osborne, and McCabe 1979) have indicated that maternal health status is related to size of deciduous and permanent dentitions. Nutrition and tooth size in living populations has not been examined. However, the role of nutrition as a contributing factor to tooth size reduction has been strongly implicated in archaeological contexts. Mark F. Guagliardo (1982) and Scott W. Simpson, Dale L. Hutchinson, and Clark Spencer Larsen (1990) have inferred that the failure of teeth to reach their maximum genetic size potential occurs in populations experiencing nutritional stress. That is, comparison of tooth size in populations dependent upon maize agriculture revealed that juveniles had consistently smaller teeth than adults. Moreover, a reduction in deciduous tooth size in comparison between hunter-gatherers and maize agriculturalists on the prehistoric southeastern U. S. coast was reported by Larsen (1983). Because



Figure I.1.9. Juvenile anterior dentition showing hypoplasias on incompletely erupted incisors. (Photograph by Barry Stark.)

deciduous tooth crowns are largely formed *in utero*, it was suggested that smaller teeth in the later period resulted from a reduction in maternal health status and placental environment.

Dental Development: Macrodefects

A final approach to assessing the nutritional and health status of contemporary and archaeological populations has to do with the analysis of enamel defects in the teeth, particularly hypoplasias (Figure I.1.9). Hypoplasias are enamel defects that typically occur as circumferential lines, grooves, or pits resulting from the death or cessation of enamel-producing cells (ameloblasts) and the failure to form enamel matrix (Goodman and Rose 1990). Goodman and Rose (1991) have reviewed a wide array of experimental, epidemiological, and bioarchaeological evidence in order to determine whether hypoplasias represent an important means for assessing nutritional status in human populations, either contemporary or archaeological. They indicate that although enamel hypoplasias arising from systemic (e.g., nutrition) versus nonsystemic factors (e.g., localized trauma) are easily identifiable, identification of an exact cause for the defects remains an intractable problem. T. W. Cutress and G. W. Suckling (1982), for example, have listed nearly 100 factors that have a causal relationship with hypoplasias, including nutritional problems. The results of a number of research projects have shown that a high frequency of individuals who have experienced malnutrition have defective enamel, thus suggesting that enamel is relatively sensitive to undernutrition. Moreover, it is a straightforward process to estimate the age at which individual hypoplasias occur based on matching the hypoplasia with dental developmental sequences (e.g., Goodman, Armelagos, and Rose 1980; Rose, Condon, and Goodman 1985; Hutchinson and Larsen 1988).

Studies based on archaeological human remains have examined hypoplasia prevalence and pattern (reviewed in Huss-Ashmore et al. 1982; Larsen 1987). In addition to determining frequency of enamel defects (which tends to be higher in agricultural populations), this research has looked at the location of defects on tooth crowns in order to examine age at the time of defect development. Contrary to earlier assertions that age pattern of defects are universal in humans, with most hypoplasias occurring in the first year of life (e.g., Sarnat and Schour 1941; see discussion in Goodman 1988), these studies have served to show that there is a great deal of variability in age of occurrence of hypoplasias. By and large, however, most reports on age patterning in hypoplasia occurrence indicate a peak in defects at 2 to 4 years of age, regardless of geographic or ecological setting (Hutchinson and Larsen 1988; Goodman and Rose 1991), a factor that most workers have attributed to nutritional stresses of postweaning diets (e.g., Corruccini, Handler, and Jacobi 1985; Webb 1989; Blakely and Mathews 1990; Simpson et al. 1990). Analyses of prevalence and pattern of hypoplasia in past populations have largely focused on recent archaeological populations. In this respect, there is a tendency for agricultural groups to show higher prevalence rates than nonagricultural (hunter-gatherer) populations (e.g., Sciulli 1978; Goodman et al. 1984a, 1984b; Hutchinson and Larsen 1988).

Unlike most other topics discussed in this chapter, this indicator of physiological stress has been investigated in ancient hominids. In the remains of early hominids in Africa, the Plio-Pleistocene australopithecines, P. V. Tobias (1967) and Tim D. White (1978) have noted the presence of hypoplasias and provided some speculation on relative health status. Of more importance, however, are the recent analyses of hypoplasias in European and Near-Eastern Nean-

derthal (Middle Paleolithic) populations. Marsha D. Ogilvie, Bryan K. Curran, and Erik Trinkaus (1989) have recorded prevalence data and estimates of developmental ages of defects on most of the extant Neanderthal teeth ($n = 669$ teeth). Their results indicate high prevalence, particularly in the permanent teeth (41.9 percent permanent teeth, 3.9 percent deciduous teeth). Although these prevalences are not as high as those observed in recent archaeological populations (e.g., Hutchinson and Larsen 1990; Van Gerven, Beck, and Hummert 1990), they do indicate elevated levels of stress in these ancient peoples. Unlike other dental series, age of occurrence of hypoplasias on the permanent dentition follows two distinct peaks, including an earlier peak between ages 2 and 5 years and a later peak between ages 11 and 13 years. The earlier peak is consistent with findings of other studies. That is, it may reflect nutritional stresses associated with weaning (Ogilvie et al. 1989). The later peak may simply represent overall high levels of systemic stress in Neanderthals. Because genetic disorders were likely eliminated from the gene pool, Ogilvie and co-workers argue against genetic agents as a likely cause. Moreover, the very low prevalence of infection in Neanderthal populations suggests that infection was an unlikely cause, leaving nutritional deficiencies, especially in the form of periodic food shortages, as the most likely causative agents.

Analysis of dental hypoplasia prevalence from specific Neanderthal sites confirms the findings of Ogilvie and co-workers, particularly with regard to the Krapina Neanderthal sample from eastern Europe (e.g., Molnar and Molnar 1985). With the Krapina dental series, Larsen and co-workers (in preparation) have made observations on the prevalence of an enamel defect known as hypocalcification, which is a disruption of the mineralization process following deposition of enamel matrix by ameloblasts. The presence of these types of enamel defects confirms the unusually high levels of stress in these early hominid populations, which is likely related to undernutrition.

Dental Development: Microdefects

An important complement to the research done on macrodefects has been observations of histological indicators of physiological stress known as Wilson bands or accentuated stria of Retzius (Rose et al. 1985; Goodman and Rose 1990). Wilson bands are features visible in thin section under low magnification ($\times 100$ to $\times 200$) as troughs or ridges in the flat enamel surface (Figure I.1.10). Concordance of these defects with hypoplasias is frequent, but certainly not universal in humans (Goodman and Rose 1990), a factor that may be related to differences in histology or etiology or both (Rose et al. 1985; Danforth 1989). K. W. Condon (1981) has concluded that Wilson bands may represent short-term stress episodes (less than one week), and hypoplasias may represent long-term stress episodes (several weeks to two months).

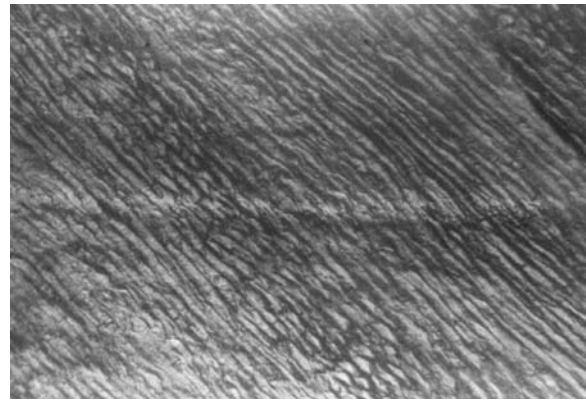


Figure I.1.10. Micrograph ($\times 160$) of canine tooth showing Wilson band from Native American Libben site. (Photograph courtesy of Jerome C. Rose.)

Jerome C. Rose, George J. Armelagos, and John W. Lallo (1978) have tested the hypothesis that as maize consumption increased and animal sources of protein consumption decreased in a weaning diet, there should be a concomitant increase in frequency of Wilson bands. Indeed, there was a fourfold increase in rate (per individual) in the full agriculturalists compared with earlier hunter-gatherers. They concluded that the declining quality of nutrition reduced the resistance of the child to infectious disease, thus increasing the individual's susceptibility to infection and likelihood of exhibiting a Wilson band. Most other studies on prehistoric populations from other cultural and geographic contexts have confirmed these findings (references cited in Rose et al. 1985).

Specific Nutritional Deficiency Diseases

Much of what is known about the nutritional quality of diets of past populations is based on the nonspecific indicators just discussed. It is important to emphasize that rarely is it possible to relate a particular hard-tissue pathology with a specific nutritional factor in archaeological human remains, not only because different nutritional problems may exhibit similar pathological signatures, but also because of the synergy between undernutrition and infection (Scrimshaw, Taylor, and Gordon 1968; Gabr 1987). This relationship has been succinctly summarized by Michael Gracey (1987: 201): "Malnourished children characteristically are enmeshed in a 'malnutrition-infection' cycle being more prone to infections which, in turn, tend to worsen the nutritional state." Thus, an episode of infection potentially exacerbates the negative effects of undernutrition as well as the severity of the pathological signature reflecting those effects.

Patricia Stuart-Macadam (1989) has reviewed evidence for the presence in antiquity of three specific nutritional diseases: scurvy, rickets, and iron-deficiency

anemia. Scurvy and rickets are produced by respective deficiencies in vitamin C (ascorbic acid) and vitamin D. Vitamin C is unusual in that it is required in the diets of humans and other primates, but only of a few other animals. Among its other functions, it serves in the synthesis of collagen, the structural protein of the connective tissues (skin, cartilage, and bone). Thus, if an individual is lacking in vitamin C, the formation of the premineralized component of bone (osteoid) will be considerably reduced.

Rickets is a disease affecting infants and young children resulting from insufficiencies in either dietary sources of vitamin D (e.g., fish and dairy products) or, of greater importance, lack of exposure to sunlight. The insufficiency reduces the ability of bone tissue to mineralize, resulting in skeletal elements (especially long bones) that are more susceptible to deformation such as abnormal bending (Figure I.1.11).

Both scurvy and rickets have been amply documented through historical accounts and in clinical settings (see Stuart-Macadam 1989). Radiographic documentation shows that bones undergoing rapid growth – namely in infants and young children – have the greatest number of changes. In infants, for example, ends of long bones and ribs are most affected and show “generalized bone atrophy and a thickening and increased density” (Stuart-Macadam 1989: 204). In children, rickets can be expressed as thin and porous bone with wide marrow spaces in general undernourishment. Alternatively, in better nourished individuals, bone tissue is more porous because of excessive bone deposition. Children with rickets can oftentimes show pronounced bowing of long bones with respect to both weight-bearing (leg) and non-weight-bearing (arm) bones.

Both scurvy and rickets, however, have been only marginally documented in the archaeological record, and mostly in historical contexts from the medieval period onward (Moller-Christiansen 1958; Maat 1982). Stuart-Macadam (1989) notes that only in the period of industrialization during the nineteenth century in Europe and North America has rickets shown an increase in prevalence.

Anemia is any condition where hemoglobin or red blood cells are reduced below normal levels. Iron-deficiency anemia is by far the most common form in living peoples, affecting more than a half billion of the current world population (Baynes and Bothwell 1990). Iron is an essential mineral, which must be ingested. It plays an important role in many body functions, especially the transport of oxygen to the body tissues (see Stuart-Macadam 1989). The bioavailability of iron from dietary sources results from several factors (see Hallberg 1981; Baynes and Bothwell 1990). With respect to its absorption, the major determinants are the sources of iron contained within foods consumed depending upon the form, heme or nonheme. Heme sources of iron from animal products are efficiently absorbed (Baynes and Bothwell 1990).

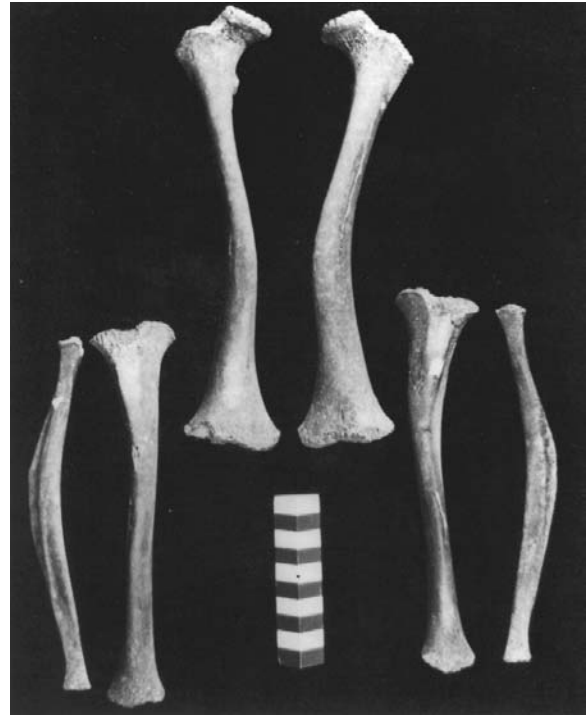


Figure I.1.11. Femora and tibiae of nineteenth-century black American showing limb bone deformation due to rickets. (Photograph by Donald J. Ortner; from Stuart-Macadam 1989, in *Reconstruction of Life from the Skeleton*, ed. Mehmet Yasar Iscan and Kenneth A. R. Kennedy; ©1989; reprinted by permission of the author and Wiley-Liss, a division of John Wiley and Sons, Inc.)

In contrast, nonheme forms of iron from the various vegetable foods have a great deal of variation in their bioavailability. Moreover, a number of substances found in foods actually inhibit iron absorption. Phytates found in many nuts (e.g., almonds, walnuts), cereals (e.g., maize, rice, whole wheat flour), and legumes inhibit dietary iron bioavailability (summarized in Baynes and Bothwell 1990). Moreover, unlike the sources of protein found in meat, plant proteins, such as soybeans, nuts, and lupines, inhibit iron absorption. Thus, populations depending on plants generally experience reduced levels of iron bioavailability. Tannates found in tea and coffee also significantly reduce iron absorption (Hallberg 1981).

There are, however, a number of foods known to enhance iron bioavailability in combination with non-heme sources of iron. For example, ascorbic acid is a very strong promotor of iron absorption (Hallberg 1981; Baynes and Bothwell 1990). Citric acid from various fruits has also been implicated in promoting iron absorption, as has lactic acid from fermented cereal beers (Baynes and Bothwell 1990). In addition, Miguel Layrisse, Carlos Martinez-Torres, and Marcel Roche

(1968; and see follow-up studies cited in Hallberg 1981) have provided experimental evidence from living human subjects that nonheme iron absorption is enhanced considerably by consumption of meat and fish, although the specific mechanism for this enhancement is not clear (Hallberg 1981).

Iron-deficiency anemia can be caused by a variety of other, nondietary factors, including parasitic infection, hemorrhage, blood loss, and diarrhea; infants can be affected by predisposing factors such as low birth weight, gender, and premature clamping of the umbilical cord (Stuart-Macadam 1989, and references cited therein). The skeletal changes observed in the clinical and laboratory settings are primarily found in the cranium, and include the following: increased width of the space between the inner and outer surfaces of the cranial vault and roof areas of the eye orbits; unusual thinning of the outer surface of the cranial vault; and a “hair-on-end” orientation of the trabecular bone between the inner and outer cranial vault (Huss-Ashmore et al. 1982; Larsen 1987; Stuart-Macadam 1989; Hill and Armelagos 1990). Postcranial changes have also been observed (e.g., Angel 1966) but are generally less severe and in reduced frequency relative to genetic anemias (Stuart-Macadam 1989). The skeletal modifications result from the hypertrophy of the blood-forming tissues in order to increase the output of red blood cells in response to the anemia (Steinbock 1976).

Skeletal changes similar to those documented in living populations have been found in archaeological human remains from virtually every region of the globe. In archaeological materials, the bony changes – pitting and/or expansion of cranial bones – have been identified by various terms, most typically *porotic hyperostosis* (Figure I.1.12) or *cribra orbitalia* (Figure I.1.13). These lesions have rarely been observed prior to the adoption of sedentism and agriculture during the Holocene, but J. Lawrence Angel (1978) has noted occasional instances extending into the Middle Pleistocene. Although the skeletal changes have been observed in individuals of all ages and both sexes, Stuart-Macadam (1985) has concluded that iron-deficiency anemia produces them in young children during the time that most of the growth in cranial bones is occurring. By contrast, the presence of porotic hyperostosis and its variants in adults represents largely anemic episodes relating to the early years of growth and development. Thus, it is not possible to evaluate iron status in adults based on this pathology.

Many workers have offered explanations for the presence of porotic hyperostosis since the pathology was first identified more than a century ago (Hill and Armelagos 1990). Recent discussions, however, have emphasized local circumstances, including nutritional deprivation brought about by focus on intensive maize consumption, or various contributing circumstances such as parasitism, diarrheal infection, or a



Figure I.1.12. Porotic hyperostosis on prehistoric Peruvian Indian posterior cranium. (Reproduced from Hrdlička 1914.)



Figure I.1.13. Cribra orbitalia in historic Florida Indian. (Photograph by Mark C. Griffin.)

combination of these factors (e.g., Hengen 1971; Carlson, Armelagos, and Van Gerven 1974; Cybulski 1977; El-Najjar 1977; Mensforth et al. 1978; Kent 1986; Walker 1986; Webb 1989). Angel (1966, 1971) argued that the primary cause for the presence of porotic hyperostosis in the eastern Mediterranean region was the presence of abnormal hemoglobins, especially thalassemia. His hypothesis, however, has remained largely unsubstantiated (Larsen 1987; Hill and Armelagos 1990).

Several human archaeological populations have been shown to have moderate to high frequencies of porotic hyperostosis after establishing agricultural economies. However, this is certainly not a ubiquitous phenomenon. For example, Larsen and co-workers (1992) and Mary L. Powell (1990) have noted that the late prehistoric populations occupying the southeastern U.S. Atlantic coast have a very low prevalence of porotic hyperostosis. These populations depended in part on maize, a foodstuff that has been implicated in reducing iron bioavailability. But a strong dependence on marine resources (especially fish) may have greatly enhanced iron absorption. In the following historic period, these native populations show marked increase in porotic hyperostosis. This probably came about because after the arrival of Europeans, consumption of maize greatly increased and that of marine resources decreased. Moreover, native populations began to use sources of water that were likely contaminated by parasites, which would have brought on an increase in the prevalence of iron-deficiency anemia (see Larsen et al. 1992).

Conclusions

This chapter has reviewed a range of skeletal and dental indicators that anthropologists have used in the reconstruction of diet and assessment of nutrition in past human populations. As noted throughout, such reconstruction and assessment, where we are dealing only with the hard-tissue remains, is especially difficult because each indicator is so often affected by other factors that are not readily controlled. For this reason, anthropologists attempt to examine as many indicators as possible in order to derive the most complete picture of diet and nutrition.

In dealing with archaeological skeletal samples, there are numerous cultural and archaeological biases that oftentimes affect the sample composition. Jane E. Buikstra and James H. Mielke have suggested:

Human groups have been remarkably creative in developing customs for disposal of the dead. Bodies have been interred, cremated, eviscerated, mummified, turned into amulets, suspended in trees, and floated down watercourses. Special cemetery areas have been reserved for persons of specific status groups or individuals who died in particular ways; for example, suicides. This variety in burial treatments can provide the archaeologist with important information about social organization in the past. On the other hand, it can also severely limit reliability of demographic parameters estimated from an excavated sample. (Buikstra and Mielke 1985:364)

Various workers have reported instances of cultural biases affecting cemetery composition. In late prehis-

toric societies in the eastern United States, young individuals and sometimes others were excluded from burial in primary cemeteries (e.g., Buikstra 1976; Russell, Choi, and Larsen 1990), although poor preservation of thinner bones of these individuals – particularly infants and young children – along with excavation biases of archaeologists, can potentially contribute to misrepresentation (Buikstra, Konigsberg, and Bullington 1986; Larsen 1987; Walker, Johnson, and Lambert 1988; Milner, Humpf, and Harpending 1989). This is not to say that skeletal samples offer a poor choice for assessing diet and nutrition in past populations. Rather, all potential biases – cultural and noncultural – must be evaluated when considering the entire record of morbidity revealed by the study of bones and teeth.

Representation in skeletal samples is made especially problematical when considering the potential for differential access to foods consumed by past human societies. For example, as revealed by analysis of prevalence of dental caries, women ate more cariogenic carbohydrates than men in many agricultural or partially agricultural societies (reviewed in Larsen 1987; Larsen et al. 1991). Even in contemporary foraging groups where food is supposedly equitably distributed among all members regardless of age or gender, various observers have found that women frequently receive less protein and fats than men, and that their diet is often nutritionally inferior to that of males (reviewed in Speth 1990). In these so-called egalitarian societies, women are regularly subject to food taboos, including a taboo on meat (e.g., Hausman and Wilmsen 1985; see discussions by Spielmann 1989 and Speth 1990). Such taboos can be detrimental especially if they are imposed during critical periods such as pregnancy or lactation (Spielmann 1989; Speth 1990). If nutritional deprivation occurs during either pregnancy or lactation, the health of the fetus or infant can be severely compromised, and delays in growth are likely to occur. Thus, when assessing nutrition in past populations, it is important that contributing factors affecting quality of diet in females or other members of societies (e.g., young children, old adults) and potential for variability in health of these individuals be carefully evaluated.

Of equal importance in the study of skeletal remains is the role of other sources of information regarding diet in archaeological settings, especially plant and animal remains. All available sources should be integrated into a larger picture, including plant and animal food remains recovered from archaeological sites, and corroborative information made available from the study of settlement patterns and ethnographic documentation of subsistence economy. The careful consideration of all these sources of information facilitates a better understanding of diet and nutrition in peoples of the past.

Clark Spencer Larsen

Bibliography

- Alvarez, J. O., J. C. Eguren, J. Caceda, and J. M. Navia. 1990. The effect of nutritional status on the age distribution of dental caries in the primary teeth. *Journal of Dental Research* 69: 1564-6.
- Alvarez, J. O., Charles A. Lewis, Carlos Saman, et al. 1988. Chronic malnutrition, dental caries, and tooth exfoliation in Peruvians aged 3-9 years. *American Journal of Clinical Nutrition* 48: 368-72.
- Alvarez, J. O., and Juan M. Navia. 1989. Nutritional status, tooth eruption, and dental caries: A review. *American Journal of Clinical Nutrition* 49: 417-26.
- Ambrose, Stanley H. 1987. Chemical and isotopic techniques of diet reconstruction in eastern North America. In *Emergent horticultural economies of the eastern Woodlands*, ed. William F. Keegan, 87-107. Center for Archaeological Investigations, Southern Illinois University Occasional Paper No. 7. Carbondale, Ill.
- Anderson, James E. 1965. Human skeletons of Tehuacan. *Science* 148: 496-7.
1967. The human skeletons. In *The prehistory of the Tehuacan Valley: Vol. 1. Environment and subsistence*, ed. Douglas S. Byers, 91-113. Austin, Texas.
- Angel, J. Lawrence. 1966. Porotic hyperostosis, anemias, malaras, and marshes in the prehistoric eastern Mediterranean. *Science* 153: 760-3.
1971. *The people of Lerna*. Washington, D.C.
1978. Porotic hyperostosis in the eastern Mediterranean. *Medical College of Virginia Quarterly* 14: 10-16.
1984. Health as a crucial factor in the changes from hunting to developed farming in the eastern Mediterranean. In *Paleopathology at the origins of agriculture*, ed. Mark Nathan Cohen and George J. Armelagos, 51-73. Orlando, Fla.
- Armelagos, George J., David S. Carlson, and Dennis P. Van Gerven. 1982. The theoretical foundations and development of skeletal biology. In *A history of American physical anthropology: 1930-1980*, ed. Frank Spencer, 305-28. New York.
- Arnaud, C. D., and S. D. Sanchez. 1990. The role of calcium in osteoporosis. *Annual Review of Nutrition* 10: 397-414.
- Atkinson, P. J. 1964. Quantitative analysis of cortical bone. *Nature* 201: 373-5.
- Aufderheide, Arthur C. 1989. Chemical analysis of skeletal remains. In *Reconstruction of life from the skeleton*, ed. Mehmet Yasar Iscan and Kenneth A. R. Kennedy, 237-60. New York.
- Bada, J. L., M. J. Schoeninger, and A. Schimmelmann. 1989. Isotopic fractionation during peptide bond hydrolysis. *Geochimica et Cosmochimica Acta* 53: 3337-41.
- Barrett, M. J., and T. Brown. 1966. Eruption of deciduous teeth in Australian aborigines. *Australian Dental Journal* 11: 43-50.
- Baynes, R. D., and T. H. Bothwell. 1990. Iron deficiency. *Annual Review of Nutrition* 10: 133-48.
- Benfer, Robert A., and Daniel S. Edwards. 1991. The principal axis method for measuring rate and amount of dental attrition: Estimating juvenile or adult tooth wear from unaged adult teeth. In *Advances in dental anthropology*, ed. Marc A. Kelley and Clark Spencer Larsen, 325-40. New York.
- Bennike, Pia. 1985. *Paleopathology of Danish skeletons*. Copenhagen.
- Blakely, Robert L., and David S. Mathews. 1990. Bioarchaeological evidence for a Spanish-Native American conflict in the sixteenth-century Southeast. *American Antiquity* 55: 718-44.
- Bogin, Barry. 1988. *Patterns of human growth*. New York.
- Boyd, Donna Catherine Markland. 1988. A functional model for masticatory-related mandibular, dental, and craniofacial microevolutionary change derived from a selected Southeastern Indian skeletal temporal series. Ph.D. thesis, University of Tennessee.
- Bray, Francesca. 1989. *The rice economies: Technology and development in Asian societies*. Cambridge, Mass.
- Brothwell, Don R. 1963. The macroscopic dental pathology of some earlier human populations. In *Dental anthropology*, ed. D. R. Brothwell, 271-88. New York.
- Brown, Antoinette B. 1988. Diet and nutritional stress. In *The King site: Continuity and contact in sixteenth century Georgia*, ed. Robert L. Blakely, 73-86. Athens, Ga.
- Buikstra, Jane E. 1976. *Hopewell in the lower Illinois River Valley: A regional approach to the study of biological variability and mortuary activity*. Northwestern University Archaeological Program, Scientific Papers No. 2.
- Buikstra, Jane E., Lyle W. Konigsberg, and Jill Bullington. 1986. Fertility and the development of agriculture in the prehistoric Midwest. *American Antiquity* 51: 528-46.
- Buikstra, Jane E., and James H. Mielke. 1985. Demography, diet, and health. In *The analysis of prehistoric diets*, ed. Robert I. Gilbert, Jr., and James H. Mielke, 359-422. Orlando, Fla.
- Bullington, Jill. 1988. Deciduous dental microwear in middle woodland and Mississippian populations from the lower Illinois River Valley. Ph.D. thesis, Northwestern University.
- Bumsted, M. Pamela. 1985. Past human behavior from bone chemical analysis: Respects and prospects. *Journal of Human Evolution* 14: 539-51.
- Burr, D. B., and R. B. Martin. 1983. The effects of composition, structure and age on the torsional properties of the human radius. *Journal of Biomechanics* 16: 603-8.
- Burr, D. B., Christopher B. Ruff, and David D. Thompson. 1990. Patterns of skeletal histologic change through time: Comparison of an archaic Native American population with modern populations. *The Anatomical Record* 226: 307-13.
- Burton, J. H., and T. D. Price. 1990. The ratio of barium to strontium as a paleodietary indicator of consumption of marine resources. *Journal of Archaeological Science* 17: 547-57.
- Calcagno, James M., and Kathleen R. Gibson. 1991. Selective compromise: Evolutionary trends and mechanisms in hominid tooth size. In *Advances in dental anthropology*, ed. Marc A. Kelley and Clark Spencer Larsen, 59-76. New York.
- Carlson, D. S., George J. Armelagos, and Dennis P. Van Gerven. 1974. Factors influencing the etiology of cribra orbitalia in prehistoric Nubia. *Journal of Human Evolution* 3: 405-10.
- Carlson, D. S., and Dennis P. Van Gerven. 1977. Masticatory function and post-Pleistocene evolution in Nubia. *American Journal of Physical Anthropology* 46: 495-506.
1979. Diffusion, biological determinism, and biocultural adaptation in the Nubian Corridor. *American Anthropologist* 81: 561-80.
- Cassidy, Claire Monod. 1984. Skeletal evidence for prehistoric subsistence adaptation in the central Ohio River Valley. In *Paleopathology at the origins of agriculture*, ed. Mark Nathan Cohen and George J. Armelagos, 307-45. Orlando, Fla.
- Cohen, Mark Nathan. 1989. *Health and the rise of civilization*. New Haven, Conn.
- Cohen, Mark Nathan, and George J. Armelagos. 1984. Paleopathology at the origins of agriculture: Editors' summa-

- tion. In *Paleopathology at the origins of agriculture*, ed. Mark Nathan Cohen and George J. Armelagos, 585-601. Orlando, Fla.
- Condon, K. W. 1981. The correspondence of developmental enamel defects between the mandibular canine and first premolar. *American Journal of Physical Anthropology* 54: 211.
- Cook, Della Collins. 1984. Subsistence and health in the lower Illinois Valley: Osteological evidence. In *Paleopathology at the origins of agriculture*, ed. Mark Nathan Cohen and George J. Armelagos, 237-69. Orlando, Fla.
1990. Epidemiology of circular caries: A perspective from prehistoric skeletons. In *A life in science: Papers in honor of J. Lawrence Angel*, ed. Jane E. Buikstra. Center for American Archeology, Scientific Papers No. 6: 64-86. Kampsville, Ill.
- Corruccini, Robert S. 1991. Anthropological aspects of orofacial and occlusal variations and anomalies. In *Advances in dental anthropology*, ed. Marc A. Kelley and Clark Spencer Larsen, 295-324. New York.
- Corruccini, Robert S., Jerome S. Handler, and Keith P. Jacobi. 1985. Chronological distribution of enamel hypoplasias and weaning in a Caribbean slave population. *Human Biology* 57: 699-711.
- Cutress, T. W., and G. W. Suckling. 1982. The assessment of non-carious defects of the enamel. *International Dental Journal* 32: 119-22.
- Cybulski, Jerome S. 1977. Cribra orbitalia, a possible sign of anemia in early historic native populations of the British Columbia coast. *American Journal of Physical Anthropology* 47: 31-40.
- Danforth, Marie Elaine. 1989. A comparison of childhood health patterns in the late classic and colonial Maya using enamel microdefects. Ph.D. thesis, Indiana University.
- El-Najjar, Mahmoud Y. 1977. Maize, malaria and the anemias in the pre-Columbian New World. *Yearbook of Physical Anthropology* 20: 329-37.
- Eveleth, Phyllis B., and James M. Tanner. 1976. *World-wide variation in human growth*. Cambridge.
- Evers, Susan E., John W. Orchard, and Richard G. Haddad. 1985. Bone density in postmenopausal North American Indian and Caucasian females. *Human Biology* 57: 719-26.
- FAO (Food and Agricultural Organization). 1970. *Amino-acid content of foods and biological data on proteins*. Rome.
- Foley, Robert, and Elizabeth Cruwys. 1986. Dental anthropology: Problems and perspectives. In *Teeth and anthropology*, ed. E. Cruwys and R. A. Foley. BAR International Series 291: 1-20. Oxford.
- Fomon, S. J., F. Haschke, E. E. Ziegler, and S. E. Nelson. 1982. Body composition of children from birth to age 10 years. *American Journal of Clinical Nutrition* 35: 1169-75.
- Freyer, David W. 1988. Caries and oral pathologies at the Mesolithic sites of Muge: Cabeço da Arruda and Moita do Sebastiao. *Trabalhos de Antropologia e Etnologia* (Portugal) 27: 9-25.
- Freeman, Leslie G. 1981. The fat of the land: Notes on Paleolithic diet in Iberia. In *Omnivorous primates: Gathering and hunting in human evolution*, ed. Robert S. O. Harding and Geza Teleki, 104-65. New York.
- Frisancho, A. Roberto. 1978. Nutritional influences on human growth and maturation. *Yearbook of Physical Anthropology* 21: 174-91.
1979. *Human adaptation: A functional interpretation*. St. Louis, Mo.
- Gabr, Mamdouh. 1987. Undernutrition and quality of life. In *Nutrition and the quality of life*, ed. G. H. Bourne. *World Review of Nutrition and Dietetics* 49: 1-21.
- Galinat, Walton C. 1985. Domestication and diffusion of maize. In *Prehistoric food production in North America*, ed. Richard I. Ford. University of Michigan, Museum of Anthropology, Anthropological Papers No. 75: 245-78. Ann Arbor.
- Garn, Stanley M., and Alphonse R. Burdi. 1971. Prenatal ordering and postnatal sequence in dental development. *Journal of Dental Research* 50: 1407-14.
- Garn, Stanley M., and D. C. Clark. 1975. Nutrition, growth, development, and maturation: Findings from the Ten-State Nutrition Survey of 1968-1970. *Pediatrics* 56: 300-19.
- Garn, Stanley, A. B. Lewis, and D. L. Polacheck. 1960. Interrelations in dental development. I. Interrelationships within the dentition. *Journal of Dental Research* 39: 1040-55.
- Garn, Stanley, R. H. Osborne, and K. D. McCabe. 1979. The effect of prenatal factors on crown dimensions. *American Journal of Physical Anthropology* 51: 665-78.
- Garn, Stanley, G. M. Owen, and D. C. Clark. 1974. Ascorbic acid: The vitamin of affluence. *Ecology of Food and Nutrition* 3: 151-3.
- Garn, Stanley, C. G. Rohmann, M. Behar, et al. 1964. Compact bone deficiency in protein-calorie malnutrition. *Science* 145: 1444-5.
- Garn, Stanley, Frederic N. Silverman, Keith P. Hertzog, and Christabel G. Rohmann. 1968. Lines and bands of increased density. *Medical Radiography and Photography* 44: 58-89.
- Goodman, Alan H. 1988. The chronology of enamel hypoplasias in an industrial population: A reappraisal of Sarnat and Schour (1941, 1942). *Human Biology* 60: 781-91.
- Goodman, Alan H., George J. Armelagos, and Jerome C. Rose. 1980. Enamel hypoplasias as indicators of stress in three prehistoric populations from Illinois. *Human Biology* 52: 512-28.
- Goodman, Alan H., and George A. Clark. 1981. Harris lines as indicators of stress in prehistoric Illinois populations. In *Biocultural adaption: Comprehensive approaches to skeletal analysis*, ed. Debra L. Martin and M. Pamela Bumsted. University of Massachusetts, Department of Anthropology, Research Report No. 20: 35-45. Amherst, Mass.
- Goodman, Alan H., John Lallo, George J. Armelagos, and Jerome C. Rose. 1984. Health changes at Dickson Mounds, Illinois (A.D. 950-1300). In *Paleopathology at the origins of agriculture*, ed. Mark Nathan Cohen and George J. Armelagos, 271-305. Orlando, Fla.
- Goodman, Alan H., Debra L. Martin, George J. Armelagos, and George Clark. 1984. Indications of stress from bones and teeth. In *Paleopathology at the origins of agriculture*, ed. Mark Nathan Cohen and George J. Armelagos, 13-49. Orlando, Fla.
- Goodman, Alan H., and Jerome C. Rose. 1990. Assessment of systemic physiological perturbations from dental enamel hypoplasias and associated histological structures. *Yearbook of Physical Anthropology* 33: 59-110.
1991. Dental enamel hypoplasias as indicators of nutritional status. In *Advances in dental anthropology*, ed. Marc A. Kelley and Clark Spencer Larsen, 279-94. New York.
- Gordon, Kathleen D. 1987. Evolutionary perspectives on human diet. In *Nutritional anthropology*, ed. Francis E. Johnston, 3-39. New York.
- Gracey, Michael. 1987. Normal growth and nutrition. In *Nutrition and the quality of life*, ed. G. H. Bourne. *World Review of Nutrition and Dietetics* 49: 160-210.
- Grine, F. E., A. J. Gwinnett, and J. H. Oaks. 1990. Early hominid dental pathology: Interproximal caries in 1.5 million-year-

- old *Paranthropus robustus* from Swartkrans. *Archives of Oral Biology* 35: 381-6.
- Grupe, Gisela, Hermann Piepenbrink, and Margaret J. Schoeninger. 1989. Note on microbial influence on stable carbon and nitrogen isotopes in bone. *Applied Geochemistry* 4: 299.
- Guagliardo, Mark F. 1982. Tooth crown size differences between age groups: A possible new indicator of stress in skeletal samples. *American Journal of Physical Anthropology* 58: 383-9.
- Hallberg, Leif. 1981. Bioavailability of dietary iron in man. *Annual Review of Nutrition* 1: 123-47.
- Harrison, G. A., J. M. Tanner, D. R. Pilbeam, and P. T. Baker. 1988. *Human biology: An introduction to human evolution, variation, growth, and adaptability*. Third edition. New York.
- Hartnady, Philip, and Jerome C. Rose. 1991. Abnormal tooth-loss patterns among Archaic-period inhabitants of the lower Pecos region, Texas. In *Advances in dental anthropology*, ed. Marc A. Kelley and Clark Spencer Larsen, 267-78. New York.
- Hausman, Alice J., and Edwin N. Wilmsen. 1985. Economic change and secular trends in the growth of San children. *Human Biology* 57: 563-71.
- Heaney, Robert P. 1993. Nutritional factors in osteoporosis. *Annual Review of Nutrition* 13: 287-316.
- Hengen, O. P. 1971. Cribra orbitalia: Pathogenesis and probable etiology. *Homo* 22: 57-75.
- Hill, M. Cassandra, and George J. Armelagos. 1990. Porotic hyperostosis in past and present perspective. In *A life in science: Papers in honor of J. Lawrence Angel*, ed. Jane E. Buikstra. Center for American Archaeology, Scientific Papers No. 6: 52-63. Kampville, Ill.
- Hillson, Simon. 1986. *Teeth*. Cambridge.
- Himes, John H. 1987. Purposeful assessment of nutritional status. In *Nutritional anthropology*, ed. Francis E. Johnston, 85-99. New York.
- Himes, John H., R. Martorell, J.-P. Habicht, et al. 1975. Patterns of cortical bone growth in moderately malnourished preschool children. *Human Biology* 47: 337-50.
- Hinton, Robert J. 1981. Form and patterning of anterior tooth wear among aboriginal human groups. *American Journal of Physical Anthropology* 54: 555-64.
- Hinton, Robert J., Maria Ostendorf Smith, and Fred H. Smith. 1980. Tooth size changes in prehistoric Tennessee Indians. *Human Biology* 52: 229-45.
- Howell, F. Clark. 1966. Observations on the earlier phases of the European Lower Paleolithic. *American Anthropologist* 68: 88-201.
1970. *Early Man*. New York.
- Howell, F. Clark, and Leslie G. Freeman. 1982. Ambrona: An Early Stone Age site on the Spanish Meseta. *L. S. B. Leakey Foundation News* No. 32: 1, 11-13.
- Hrdlička, Aleš. 1914. *Anthropological work in Peru in 1913, with notes on the pathology of the ancient Peruvians*. Smithsonian Miscellaneous Collections 61, No. 18. Washington, D.C.
- Hummert, James R., and Dennis P. Van Gerven. 1983. Skeletal growth in a medieval population from Sudanese Nubia. *American Journal of Physical Anthropology* 60: 471-8.
- Huss-Ashmore, Rebecca, Alan H. Goodman, and George J. Armelagos. 1982. Nutritional inference from paleopathology. *Advances in Archeological Method and Theory* 5: 395-474.
- Hutchinson, Dale L., and Clark Spencer Larsen. 1988. Stress and lifeway change: The evidence from enamel hypoplasias. In *The archaeology of mission Santa Catalina de Guale: 2. Biocultural interpretations of a population in transition*, ed. Clark Spencer Larsen. Anthropological Papers of the American Museum of Natural History No. 68: 50-65. New York.
- Inoue, N., G. Ito, and T. Kamegai. 1986. Dental pathology of hunter-gatherers and early farmers in prehistoric Japan. In *Prehistoric hunter gatherers in Japan: New research methods*, ed. Takeru Akazawa and C. Melvin Aikens. University of Tokyo, University Museum, Bulletin No. 27: 163-98. Tokyo.
- Jantz, Richard L., and Douglas W. Owsley. 1984. Long bone growth variation among Arikara skeletal populations. *American Journal of Physical Anthropology* 63: 13-20.
- Johnston, Francis E. 1986. Somatic growth of the infant and preschool child. In *Human growth*, Vol. 2, ed. F. Falkner and J. M. Tanner, 3-24. New York.
- Jones, P. R. M., and R. F. A. Dean. 1959. The effects of kwashiorkor on the development of the bones of the knee. *Journal of Pediatrics* 54: 176-84.
- Katz, Solomon H. 1987. Food and biocultural evolution: A model for the investigation of modern nutritional problems. In *Nutritional anthropology*, ed. Francis E. Johnston, 41-63. New York.
- Katz, Solomon H., M. L. Hediger, and L. A. Valleroy. 1974. Traditional maize processing techniques in the New World. *Science* 184: 765-73.
1975. The anthropological and nutritional significance of traditional maize processing techniques in the New World. In *Biosocial interrelations in population adaptation*, ed. Elizabeth S. Watts, Francis E. Johnston, and Gabriel Lasker, 195-231. The Hague.
- Keegan, William F. 1989. Stable isotope analysis of prehistoric diet. In *Reconstruction of life from the skeleton*, ed. Mehmet Yasar Iscan and Kenneth A. R. Kennedy, 223-36. New York.
- Kelley, Marc A., Dianne R. Levesque, and Eric Weidl. 1991. Contrasting patterns of dental disease in five early northern Chilean groups. In *Advances in dental anthropology*, ed. Marc A. Kelley and Clark Spencer Larsen, 203-13. New York.
- Kennedy, Kenneth A. R. 1984. Growth, nutrition, and pathology in changing paleodemographic settings in South Asia. In *Paleopathology at the origins of agriculture*, ed. Mark Nathan Cohen and George J. Armelagos, 169-92. Orlando, Fla.
- Kent, Susan. 1986. The influence of sedentism and aggregation on porotic hyperostosis and anaemia: A case study. *Man* 21: 605-36.
- Kimura, Kunihiro. 1984. Studies on growth and development in Japan. *Yearbook of Physical Anthropology* 27: 179-213.
- Klein, Richard G. 1987. Problems and prospects in understanding how early people exploited animals. In *The evolution of human hunting*, ed. M. H. Netecki and D. V. Netecki, 11-45. New York.
1989. *The human career: Human biological and cultural origins*. Chicago.
- Klepinger, Linda L. 1984. Nutritional assessment from bone. *Annual Review of Anthropology* 13: 75-96.
- Krouse, R. H. 1987. Sulphur and carbon isotope studies of food chains. In *Diet and subsistence: Current archaeological perspectives*, ed. B. V. Kennedy and G. M. LeMoine. University of Calgary Archaeological Association, Proceedings of the 19th Annual Conference. Calgary, Alberta.
- Lallo, John W. 1973. The skeletal biology of three prehistoric American Indian societies from the Dickson mounds. Ph.D. thesis, University of Massachusetts.
- Larsen, Clark Spencer. 1982. *The anthropology of St. Catherines Island: 1. Prehistoric human biological adapta-*

- tion. Anthropological Papers of the American Museum of Natural History No. 57 (part 3). New York.
1983. Deciduous tooth size and subsistence change in prehistoric Georgia coast populations. *Current Anthropology* 24: 225-6.
1987. Bioarchaeological interpretations of subsistence economy and behavior from human skeletal remains. *Advances in Archaeological Method and Theory* 10: 339-445.
- Larsen, Clark Spencer, Christopher B. Ruff, Margaret J. Schoeninger, and Dale L. Hutchinson. 1992. Population decline and extinction in La Florida. In *Disease and demography in the Americas: Changing patterns before and after 1492*, ed. Douglas H. Ubelaker and John W. Verano. Washington, D.C.
- Larsen, Clark Spencer, Margaret J. Schoeninger, Dale L. Hutchinson, et al. 1990. Beyond demographic collapse: Biological adaptation and change in native populations of La Florida. In *Columbian consequences, Vol 2: Archaeological and historical perspectives on the Spanish borderlands east*, ed. David Hurst Thomas, 409-28. Washington, D.C.
- Larsen, Clark Spencer, Rebecca Shavit, and Mark C. Griffin. 1991. Dental caries evidence for dietary change: An archaeological context. In *Advances in dental anthropology*, ed. Marc A. Kelley and Clark Spencer Larsen, 179-202. New York.
- Layrisse, Miguel, Carlos Martinez-Torres, and Marcel Roche. 1968. Effect of interaction of various foods on iron absorption. *American Journal of Clinical Nutrition* 21: 1175-83.
- Lindsay, Robert, and Felicia Cosman. 1992. Primary osteoporosis. In *Disorders of bone and primary metabolism*, ed. Frederic L. Coe and Murray J. Favus, 831-88. New York.
- Lovejoy, C. Owen, Katherine F. Russell, and Mary L. Harrison. 1990. Long bone growth velocity in the Libben population. *American Journal of Human Biology* 2: 533-41.
- Maat, G. J. R. 1982. Scurvy in Dutch whalers buried at Spitsbergen. *Paleopathology Association, Papers on Paleopathology, Fourth European Members Meeting*, ed. Eve Cockburn, 8. Detroit, Mich.
- Malina, Robert M. 1987. Nutrition and growth. In *Nutritional anthropology*, ed. Francis E. Johnston, 173-96. New York.
- Mann, Alan E. 1981. Diet and human evolution. In *Omnivorous primates: Gathering and hunting in human evolution*, ed. Robert S. O. Harding and Geza Teleki, 10-36. New York.
- Martin, Debra L., and George J. Armelagos. 1985. Skeletal remodeling and mineralization as indicators of health: An example from prehistoric Sudanese Nubia. *Journal of Human Evolution* 14: 527-37.
- Martin, Debra L., Alan H. Goodman, and George J. Armelagos. 1985. Skeletal pathologies as indicators of quality and quantity of diet. In *The analysis of prehistoric diets*, ed. Robert I. Gilbert, Jr., and James H. Mielke, 227-79. Orlando, Fla.
- McElroy, Ann, and Patricia K. Townsend. 1979. *Medical Anthropology*. Belmont, Calif.
- McGregor, I. A., A. M. Thomson, and W. Z. Billewicz. 1968. The development of primary teeth in children from a group of Gambian villages and critical examination of its use for estimating age. *British Journal of Nutrition* 22: 307-14.
- McHenry, Henry. 1968. Transverse lines in long bones of prehistoric California Indians. *American Journal of Physical Anthropology* 29: 1-18.
- McLaren, Donald S. 1976. *Nutrition in the community*. New York.
- Meiklejohn, Christopher, Catherine Schentag, Alexandra Venema, and Patrick Key. 1984. Socioeconomic change and patterns of pathology and variation in the Mesolithic and Neolithic of western Europe: Some suggestions. In *Paleopathology at the origins of agriculture*, ed. Mark Nathan Cohen and George J. Armelagos, 75-100. Orlando, Fla.
- Mensforth, Robert P., C. Owen Lovejoy, John W. Lallo, and George J. Armelagos. 1978. The role of constitutional factors, diet, and infectious disease in the etiology of porotic hyperostosis and periosteal reactions in prehistoric infants and children. *Medical Anthropology* 2: 1-59.
- Merchant, Virginia L., and Douglas H. Ubelaker. 1977. Skeletal growth of the protohistoric Arikara. *American Journal of Physical Anthropology* 46: 61-72.
- Milner, George R. 1984. Dental caries in the permanent dentition of a Mississippian period population from the American Midwest. *Collegium Antropologicum* 8: 77-91.
- Milner, George R., Dorothy A. Humpf, and Henry C. Harpending. 1989. Pattern matching of age-at-death distributions in paleodemographic analysis. *American Journal of Physical Anthropology* 80: 49-58.
- Moller-Christiansen, V. 1958. *Bogen om Aebelbolt Kloster*. Copenhagen.
- Molnar, Stephen. 1972. Tooth wear and culture: A survey of tooth functions among some prehistoric populations. *Current Anthropology* 13: 511-26.
- Molnar, Stephen, and I. M. Molnar. 1985. The incidence of enamel hypoplasia among the Krapina Neanderthals. *American Anthropologist* 87: 536-49.
- Moore, K. P., S. Thorp, and D. P. Van Gerven. 1986. Pattern of dental eruption, skeletal maturation and stress in a medieval population from Sudanese Nubia. *Human Evolution* 1: 325-30.
- Mottram, R. F. 1979. *Human nutrition*. Third edition. Westport, Conn.
- Newbrun, Ernest. 1982. Sugar and dental caries: A review of human studies. *Science* 217: 418-23.
- Nordin, B. E. C. 1973. *Metabolic bone and stone disease*. Baltimore, Md.
- Ogilvie, Marsha D., Bryan K. Curran, and Erik Trinkaus. 1989. Incidence and patterning of dental enamel hypoplasia among the Neanderthals. *American Journal of Physical Anthropology* 79: 25-41.
- Ortner, Donald J., and Walter G. J. Putschar. 1985. *Identification of pathological conditions in the human skeletal remains*. Washington, D.C.
- Owsley, Douglas W., and Richard L. Jantz. 1985. Long bone lengths and gestational age distributions of post-contact period Arikara Indian perinatal infant skeletons. *American Journal of Physical Anthropology* 68: 321-8.
- Patterson, David Kingsnorth, Jr. 1984. *A diachronic study of dental palaeopathology and attritional status of prehistoric Ontario pre-Iroquois and Iroquois populations*. Archaeological Survey of Canada, Paper No. 122. Ottawa.
- Powell, Mary Lucas. 1985. The analysis of dental wear and caries for dietary reconstruction. In *The analysis of prehistoric diets*, ed. Robert I. Gilbert, Jr., and James H. Mielke, 307-38. Orlando, Fla.
1990. On the eve of the conquest: Life and death at Irene Mound, Georgia. In *The archaeology of mission Santa Catalina de Guale: 2. Biocultural interpretations of a population in transition*, ed. Clark Spencer Larsen. Anthropological Papers of the American Museum of Natural History No. 68: 26-35. New York.
- Price, T. Douglas. 1985. Late archaic subsistence in the mid-western United States. *Journal of Human Evolution* 14: 449-59.

- Roe, Daphne A. 1973. *A plague of corn: The social history of pellagra*. Ithaca, N.Y.
- Rose, Jerome C., George J. Armelagos, and John W. Lallo. 1978. Histological enamel indicator of childhood stress in prehistoric skeletal samples. *American Journal of Physical Anthropology* 49: 511-16.
- Rose, Jerome C., Barbara A. Burnett, Michael S. Nassaney, and Mark W. Blauer. 1984. Paleopathology and the origins of maize agriculture in the lower Mississippi Valley and Caddoan areas. In *Paleopathology at the origins of agriculture*, ed. Mark Nathan Cohen and George J. Armelagos, 393-424. Orlando, Fla.
- Rose, Jerome C., Keith W. Condon, and Alan H. Goodman. 1985. Diet and dentition: Developmental disturbances. In *The analysis of prehistoric diets*, ed. Robert I. Gilbert, Jr., and J. H. Mielke, 281-305. Orlando, Fla.
- Rose, Jerome C., Murray K. Marks, and Larry L. Tieszen. 1991. Bioarchaeology and subsistence in the central and lower portions of the Mississippi Valley. In *What mean these bones? Integrated studies in southeastern bioarchaeology*, ed. Mary Lucas Powell, Patricia S. Bridges, and Ann Marie Mires, 7-21. Tuscaloosa, Ala.
- Ross, Harold M. 1976. Bush fallow farming, diet and nutrition: A Melanesian example of successful adaptation. In *The measures of man: Methodologies in biological anthropology*, ed. Eugene Giles and Jonathan S. Friedlaender, 550-615. Cambridge, Mass.
- Ruff, Christopher. 1991. Biomechanical analyses of archaeological human skeletal samples. In *The skeletal biology of past peoples: Advances in research methods*, ed. Shelley R. Saunders and Mary A. Katzenberg. New York.
- Ruff, Christopher, and Wilson C. Hayes. 1982. Subperiosteal expansion and cortical remodeling of the human femur and tibia with aging. *Science* 217: 945-8.
- Ruff, Christopher, and Clark Spencer Larsen. 1990. Postcranial biomechanical adaptations to subsistence strategy changes on the Georgia coast. In *The archaeology of mission Santa Catalina de Guale: 2. Biocultural interpretations of a population in transition*, ed. Clark Spencer Larsen. Anthropological Papers of the American Museum of Natural History No. 68: 94-120. New York.
- Russell, Katherine F., Inui Choi, and Clark Spencer Larsen. 1990. The paleodemography of Santa Catalina de Guale. In *The archaeology of mission Santa Catalina de Guale: 2. Biocultural interpretations of a population in transition*, ed. Clark Spencer Larsen. Anthropological Papers of the American Museum of Natural History No. 68: 36-49. New York.
- Sandford, Mary K., ed. 1993. *Investigations of ancient human tissue: Chemical analyses in anthropology*. Langhorne, Pa.
- Sarnat, B. G., and I. Schour. 1941. Enamel hypoplasias (chronic enamel aplasia) in relationship to systemic diseases: A chronologic, morphologic, and etiologic classification. *Journal of the American Dental Association* 28: 1989-2000.
- Schaafsma, G., E. C. H. van Beresteyn, J. A. Raymakers, and S. A. Duursma. 1987. Nutritional aspects of osteoporosis. In *Nutrition and the quality of life*, ed. G. H. Bourne. *World Review of Nutrition and Dietetics* 49: 121-59.
- Schoeninger, Margaret J. 1989. Reconstructing prehistoric human diet. In *Chemistry of prehistoric human bone*, ed. T. Douglas Price, 38-67. New York.
- Schoeninger, Margaret J., and Michael J. DeNiro. 1984. Nitrogen and carbon isotopic composition of bone collagen from marine and terrestrial animals. *Geochimica et Cosmochimica Acta* 48: 625-39.
- Schoeninger, Margaret J., Katherine M. Moore, Matthew L. Murray, and John D. Kingston. 1989. Detection of bone preservation in archaeological and fossil samples. *Applied Geochemistry* 4: 281-92.
- Schoeninger, Margaret J., and C. S. Peebles. 1981. Effects of mollusk eating on human bone strontium levels. *Journal of Archaeological Science* 8: 391-7.
- Schoeninger, Margaret J., Nikolaas J. van der Merwe, and Katherine Moore. 1990. Decrease in diet quality between the prehistoric and contact periods. In *The archaeology of mission Santa Catalina de Guale: 2. Biocultural interpretations of a population in transition*, ed. Clark Spencer Larsen. Anthropological Papers of the American Museum of Natural History No. 68: 78-93. New York.
- Sciulli, Paul W. 1978. Developmental abnormalities of the permanent dentition in prehistoric Ohio Valley Amerindians. *American Journal of Physical Anthropology* 48: 193-8.
- Scott, G. Richard, and Christy G. Turner II. 1988. Dental anthropology. *Annual Review of Anthropology* 17: 99-126.
- Scrimshaw, N. S., C. E. Taylor, and J. E. Gordon. 1968. *Interaction of nutrition and infection*. World Health Organization Monograph 57. Geneva.
- Sealy, Judith. 1986. *Stable carbon isotopes and prehistoric diets in the south-western Cape Province, South Africa*. Cambridge Monographs in African Archaeology 15, BAR International Series 293. Oxford.
- Shipman, Pat. 1986a. Scavenging or hunting in early hominids: Theoretical framework and tests. *American Anthropologist* 88: 27-43.
- 1986b. Studies of hominid-faunal interactions at Olduvai Gorge. *Journal of Human Evolution* 15: 691-706.
- Shipman, Pat, Wendy Bosler, and Karen Lee Davis. 1981. Butchering of giant geladas at an Acheulean site. *Current Anthropology* 22: 257-68.
- Sillen, Andrew, and Maureen Kavanaugh. 1982. Strontium and paleodietary research: A review. *Yearbook of Physical Anthropology* 25: 67-90.
- Simpson, Scott W., Dale L. Hutchinson, and Clark Spencer Larsen. 1990. Coping with stress: Tooth size, dental defects, and age-at-death. In *The archaeology of the mission Santa Catalina de Guale: 2. Biocultural interpretations of a population in transition*, ed. Clark Spencer Larsen. Anthropological Papers of the American Museum of Natural History No. 68: 66-77. New York.
- Smith, B. Holly. 1984. Patterns of molar wear in hunter-gatherers and agriculturalists. *American Journal of Physical Anthropology* 63: 39-56.
1985. Development and evolution of the helicoidal plane of dental occlusion. *American Journal of Physical Anthropology* 69: 21-35.
1991. Standards of human tooth formation and dental age assessment. In *Advances in dental anthropology*, ed. Marc A. Kelley and Clark Spencer Larsen, 143-68. New York.
- Smith, Bruce D. 1989. Origins of agriculture in eastern North America. *Science* 246: 1566-72.
- Smith, Fred H., Maria Ostendorf Smith, and Robert J. Hinton. 1980. Evolution of tooth size in the prehistoric inhabitants of the Tennessee Valley. In *The skeletal biology of aboriginal populations in the southeastern United States*, ed. P. Willey and F. H. Smith. Tennessee Anthropological Association Miscellaneous Paper No. 5: 81-103. Knoxville, Tenn.
- Smith, Patricia, Ofer Bar-Yosef, and Andrew Sillen. 1984. Archaeological and skeletal evidence for dietary change during the late Pleistocene/early Holocene in the Levant. In *Paleopathology at the origins of agriculture*, ed. Mark Nathan Cohen and George J. Armelagos, 101-36. Orlando, Fla.
- Speth, John D. 1990. Seasonality, resource stress, and food

- sharing in so-called "egalitarian" foraging societies. *Journal of Anthropological Archaeology* 9: 148-88.
- Spielmann, Katherine A. 1989. A review: Dietary restrictions on hunter-gatherer women and the implications for fertility and infant mortality. *Human Ecology* 17: 321-45.
- Stafford, Thomas W., Klaus Brendel, and Raymond C. Duhamel. 1988. Radiocarbon, ^{13}C and ^{15}N analysis of fossil bone: Removal of humates with XAD-2 resin. *Geochimica et Cosmochimica Acta* 52: 2257-67.
- Steinbock, R. Ted. 1976. *Paleopathological diagnosis and interpretation*. Springfield, Ill.
- Stewart, R. J. C. 1975. Bone pathology in experimental malnutrition. *World Review of Nutrition and Dietetics* 21: 1-74.
- Stewart, R. J. C., and B. S. Platt. 1958. Arrested growth lines in the bones of pigs on low-protein diets. *Proceedings of the Nutrition Society* 17: v-vi.
- Stini, William A. 1971. Evolutionary implications of changing nutritional patterns in human populations. *American Anthropologist* 73: 1019-30.
1990. "Osteoporosis": Etiologies, prevention, and treatment. *Yearbook of Physical Anthropology* 33: 151-94.
- Stout, Samuel D. 1978. Histological structure and its preservation in ancient bone. *Current Anthropology* 19: 601-3.
1983. The application of histomorphometric analysis to ancient skeletal remains. *Anthropos (Greece)* 10: 60-71.
1989. Histomorphometric analysis of human skeletal remains. In *Reconstruction of life from the skeleton*, ed. Mehmet Yasar Iscan and Kenneth A. R. Kennedy, 41-52. New York.
- Stout, Samuel D., and David J. Simmons. 1979. Use of histology in ancient bone research. *Yearbook of Physical Anthropology* 22: 228-49.
- Stout, Samuel D., and Steven L. Teitelbaum. 1976. Histological analysis of undercalcified thin sections of archaeological bone. *American Journal of Physical Anthropology* 44: 263-70.
- Stuart-Macadam, Patricia. 1985. Porotic hyperostosis: Representative of a childhood condition. *American Journal of Physical Anthropology* 66: 391-8.
1989. Nutritional deficiency diseases: A survey of scurvy, rickets, and iron-deficiency anemia. In *Reconstruction of life from the skeleton*, ed. Mehmet Yasar Iscan and Kenneth A. R. Kennedy, 201-22. New York.
- Sundick, Robert I. 1978. Human skeletal growth and age determination. *Homo* 29: 228-49.
- Teaford, Mark F. 1991. Dental microwear: What can it tell us about diet and dental function? In *Advances in dental anthropology*, ed. Marc A. Kelley and Clark Spencer Larsen, 341-56. New York.
- Thompson, D. D. 1980. Age changes in bone mineralization, cortical thickness, and Haversian canal area. *Calcified Tissue International* 31: 5-11.
- Tobias, P. V. 1967. *The cranium and maxillary dentition of zinjantropus (australopithecus) boisei*. Cambridge.
- Turner, Christy G., II. 1979. Dental anthropological indications of agriculture among the Jomon people of central Japan. *American Journal of Physical Anthropology* 51: 619-36.
- Turner, Christy G., II, and Lilia M. Cheuiche Machado. 1983. A new dental wear pattern and evidence for high carbohydrate consumption in a Brazilian Archaic skeletal population. *American Journal of Physical Anthropology* 61: 125-30.
- Van der Merwe, Nikolaas J. 1982. Carbon isotopes, photosynthesis and archaeology. *American Scientist* 70: 596-606.
- Van Gerven, Dennis P., Rosemary Beck, and James R. Humer. 1990. Patterns of enamel hypoplasia in two medieval populations from Nubia's Batn El Hajar. *American Journal of Physical Anthropology* 82: 413-20.
- Walker, Phillip L. 1978. A quantitative analysis of dental attrition rates in the Santa Barbara Channel area. *American Journal of Physical Anthropology* 48: 101-6.
1986. Porotic hyperostosis in a marine-dependent California Indian population. *American Journal of Physical Anthropology* 69: 345-54.
- Walker, Phillip L., Gregory Dean, and Perry Shapiro. 1991. Estimating age from tooth wear in archaeological populations. In *Advances in dental anthropology*, ed. Marc A. Kelley and Clark Spencer Larsen, 169-78. New York.
- Walker, Phillip L., and Jon M. Erlandson. 1986. Dental evidence for prehistoric dietary change on northern Channel Islands, California. *American Antiquity* 51: 375-83.
- Walker, Phillip L., John Johnson, and Patricia Lambert. 1988. Age and sex biases in the preservation of human skeletal remains. *American Journal of Physical Anthropology* 76: 183-8.
- Wallace, John A. 1974. Approximal grooving of teeth. *American Journal of Physical Anthropology* 40: 385-90.
- Webb, Stephen. 1989. *Prehistoric stress in Australian aborigines: A paleopathological study of a hunter-gatherer population*. International Series 490. Oxford.
- White, Tim D. 1978. Early hominid enamel hypoplasia. *American Journal of Physical Anthropology* 49: 79-84.
- Wing, Elizabeth S., and Antoinette B. Brown. 1979. *Paleonutrition*. New York.
- Wolf, George. 1980. Vitamin A. In *Human nutrition*, ed. R. B. Alfin-Slater and D. Kritchevsky, 97-203. New York.
- Wu Xinzhi and Zhang Zhenbiao. 1985. *Homo sapiens* remains from late palaeolithic and neolithic China. In *Paleo-anthropology and palaeolithic archaeology in the People's Republic of China*, ed. Wu Rukang and John W. Olsen, 107-33. Orlando, Fla.
- Yagi, Tamotsu, Yoshihide Takebe, and Minoru Itoh. 1989. Secular trends in physique and physical fitness in Japanese students during the last 20 years. *American Journal of Human Biology* 1: 581-7.
- Yarbrough, C., J.-P. Habicht, R. Martorell, and R. E. Klein. 1974. Anthropometry as an index of nutritional status. In *Nutrition and malnutrition: Identification and measurement*, ed. A. F. Roche and F. Falkner, 15-26. New York.
- y'Edynak, Gloria. 1989. Yugoslav Mesolithic dental reduction. *American Journal of Physical Anthropology* 78: 17-36.
- y'Edynak, Gloria, and Sylvia Fleisch. 1983. Microevolution and biological adaptability in the transition from food-collecting to food-producing in the Iron Gates of Yugoslavia. *Journal of Human Evolution* 12: 279-96.

I.2. 🦴 Paleopathological Evidence of Malnutrition

The quantity and nutritional quality of food available to human populations undoubtedly played a major role in the adaptive processes associated with human evolution. This should have been particularly the case in that period of human history from Mesolithic times to the present when epochal changes took place in the subsistence base of many human societies. In the Near East

the domestication of plants and animals began toward the end of the Mesolithic period but became fully developed in the Neolithic. This development included agriculture and pastoralism along with cultural changes associated with greater sedentism and urbanism.

Paleopathology, primarily through the study of human skeletal remains, has attempted to interpret the impact such changes have had upon human health. A recent focus has been on the transition from a hunting and gathering way of life to one associated with incipient or fully developed agriculture (e.g., Cohen and Armelagos 1984b; Cohen 1989; Meiklejohn and Zvelebil 1991). One of the questions being asked is whether greater dependence on fewer food sources increased human vulnerability to famine and malnutrition. The later transition into an increasingly sedentary urban existence in the Bronze and Iron Ages has not been as carefully studied. However, analysis of data from skeletal remains in numerous archaeological sites is providing insight into some of the effects upon nutrition that increasing human density and attendant subsistence changes have had.

In the study of prehistoric health, perhaps the least complex nutritional data comes from human remains that have been mummified. Preservation of human soft tissues occurs either naturally, as in the bogs of northern Europe and very arid areas of the world, or through cultural intervention with embalming methods. Some mummies have provided direct evidence of diet from the intestinal contents of their stomachs (e.g., Glob 1971: 42-3; Fischer 1980: 185-9; Brothwell 1986: 92). However, the most ubiquitous source of data comes from human skeletal remains where the impact of dietary factors tends to be indirect, limited, and difficult to interpret.

Generally, only about 10 percent of a typical sample of human archaeological burials will show any significant evidence of skeletal disease. (Clearly the people represented by the normal-appearing burials died of something, but there are no anatomical features that help determine what this might have been.) Of the 10 percent showing pathology, about 90 percent of their disease conditions resulted from trauma, infection, or arthritis – the three predominant pathological skeletal conditions. All other diseases, including those that might be caused by malnutrition, are incorporated in the residual 10 percent, meaning that even in a large sample of archaeological skeletons, one is unlikely to find more than a few examples of conditions that might be attributable to nutritional problems.

Once a pathological condition due to malnutrition is recognized in bone, correct diagnosis is challenging. Identification begins with those nutritional diseases most commonly known today that can affect the skeleton. These are: (1) vitamin D deficiency, (2) vitamin C deficiency, (3) iodine deficiency, (4) iron deficiency, (5) excessive dietary fluorine, (6) protein-calorie deficiency, and (7) trace element deficiencies.

Care needs to be exercised both in establishing a

preferred diagnosis for a pathological condition and in interpreting the diagnoses of others. This is particularly the case in interpreting evidence of malnutrition in archaeological remains. Malnutrition is a general state that may cause more than one pathological condition in the same individual, and it may also be accompanied by other disease conditions, making the pathological profile complex and confusing. For example, scurvy and rickets may appear together (Follis, Jackson, and Park 1940), or scurvy may be associated with iron-deficiency anemia (Goldberg 1963).

All of these issues place significant limitations on reconstructing nutritional problems in antiquity. However, emerging research methods, such as stable isotope analysis and bone histology, evidence from related fields, such as dental pathology, and new areas of research concentration, such as infant skeletal studies, may provide additional data. Analysis of stable isotopes of human bone collagen allows us to determine the balance of food sources between terrestrial animal, marine animal, and plant materials (Katzenberg 1992). Isotope analysis of human hair may provide a more refined breakdown of plant materials eaten over a much shorter period than the 25 to 30 years that bone collagen analysis provides (White 1993: 657). C. D. White (1993: 657), working on prehistoric human remains from Nubia, has claimed that isotopic analysis of hair points to a seasonal difference between consumption of plants such as wheat, barley, and most fruits and vegetables and consumption of the less nutritious plants such as sorghum and millet.

Analysis of bone histology by M. Schultz and fellow workers (Schultz 1986, 1990, 1993; Carli-Thiele and Schultz 1994; Schultz and Schmidt-Schultz 1994) has identified features that assist in differential diagnosis in archaeological human skeletal remains. In one study of human remains from an Early Bronze Age (2500 to 2300 B.C.) cemetery in Anatolia, Schultz (1993: 189) detected no anatomical evidence of rickets in an infant sample. However, microscopic examination revealed a rickets prevalence of 4 percent.

Dental paleopathology provides an additional dimension to understanding nutritional problems. For example, caries rate and location may help identify what type of food was eaten (e.g., Littleton and Frohlich 1989, 1993; Meiklejohn and Zvelebil 1991). Enamel hypoplasias, which are observable defects in dental enamel, may provide information about timing and severity of nutritional stress (Goodman, Martin, and Armelagos 1984; Goodman 1991; Meiklejohn and Zvelebil 1991). Patterns of antemortem tooth loss may suggest whether an individual suffered from a nutritional disease, such as scurvy (Maat 1986: 158), or excess calculus or poor dental hygiene (Lukacs 1989).

Thorough analysis of skeletons of infants and children, which until recently has received minimal attention, can also provide valuable information on the health of a population. Indeed, because of a child's rapid growth and consequent need for optimal nutrition,

immature skeletons will reflect the nutritional status of a population better than those of adults. This is especially the case with diseases such as scurvy, rickets, and iron-deficiency anemia, whose impact is greatest on children between the ages of 6 months and 2 years (Stuart-Macadam 1989a:219).

In this chapter, we discuss skeletal abnormalities associated with nutritional diseases for which there is archaeological skeletal evidence in various geographical areas and time periods in the Old World. These diseases are: vitamin D deficiency, vitamin C deficiency, iron deficiency, fluorosis, and protein-calorie deficiency. We will focus on anatomical evidence of nutritional disease but will include other types of evidence as it occurs. For a discussion of the pathogenesis of these diseases we refer the reader to other sources (Ortner and Putschar 1981; Resnick and Niwayama 1988).

Vitamin D Deficiency

Vitamin D deficiency causes rickets in children and osteomalacia in adults. In general, these conditions should be rare in societies where exposure to sunlight is common, as the body can synthesize vitamin D precursors with adequate sunlight. In fact, there has been some speculation that rickets will not occur in areas of abundant sunlight (Angel 1971: 89). Cultural factors, however, may intervene. The use of concealing clothing such as veils, the practice of long-term sequestration of women (*purdah*), or the swaddling of infants (Kuhnke 1993: 461) will hinder the synthesis of vitamin D. Thus, in modern Asia both rickets and osteomalacia have been reported, with the condition attributed to culturally patterned avoidance of sunlight (Fallon 1988: 1994). In the Near East and North Africa cases of rickets have been reported in large towns and sunless slums (Kuhnke 1993: 461).

Vitamin D is critical to the mineralization of bone protein matrix. If the vitamin is not present during bone formation, the protein matrix does not mineralize. Turnover of bone tissue is most rapid during the growth phase, and in rickets much of the newly forming protein matrix may not be mineralized. This compromises biomechanical strength; bone deformity may occur, especially in the weight-bearing limbs, and may be apparent in archaeological human remains.

In the active child, the deformity tends to be in the extremities, and its location may be an indication of when the individual suffered from this disease. Deformities that are restricted to the upper limbs may indicate that the child could not yet walk (Ortner and Putschar 1981: 278), whereas those that show bowing of both the upper and lower limbs may be indicative of chronic or recurring rickets (Stuart-Macadam 1989b: 41). Bowing limited to the long bones of the lower extremities would indicate that rickets had become active only after the child had started walking (Ortner and Putschar 1981: 278).

There is a relatively rare form of rickets that is not caused by a deficiency in dietary vitamin D. Instead, this condition results from the kidneys' failure to retain phosphorus (Fallon 1988: 1994), and as phosphate is the other major component of bone mineral besides calcium, the effect is deficient mineralization as well. This failure may be caused by a congenital defect in the kidneys or by other diseases affecting the kidneys. The importance of nondietary rickets to this chapter is that the anatomical manifestations in the skeleton are indistinguishable from those caused by vitamin D deficiency.

The adult counterpart of rickets in the skeletal record is osteomalacia, whose expression requires an even more severe state of malnutrition (Maat 1986: 157). Women are vulnerable to osteomalacia during pregnancy and lactation because their need for calcium is great. If dietary calcium is deficient, the developing fetus will draw on calcium from the mother's skeleton. If vitamin D is also deficient, replacement of the mineral used during this period will be inhibited even if dietary calcium becomes available. As in rickets, biomechanical strength of bone may be inadequate, leading to deformity. This deformity is commonly expressed in the pelvis as biomechanical forces from the femoral head compress the anteroposterior size of the pelvis and push the acetabula into the pelvic canal.

Undisputed anatomical evidence of rickets or osteomalacia in archaeological remains is uncommon for several reasons. First, criteria for diagnosis of these conditions in dry bone specimens have not been clearly distinguished from some skeletal manifestations of other nutritional diseases such as scurvy or anemia. Second, reports on cases of rickets are often based on fairly subtle changes in the shape of the long bones (Bennike 1985: 210, 213; Grmek 1989: 76), which may not be specific for this condition. Third, cases of rickets that are associated with undernourishment are difficult to recognize because growth may have stopped (Stuart-Macadam 1989b: 41).

A remarkable case from a pre-Dynastic Nubian site illustrates the complexity of diagnosis in archaeological human remains. The case has been described by J. T. Rowling (1967: 277) and by D. J. Ortner and W. G. J. Putschar (1981: 284-7). The specimen exhibits bending of the long bones of the forearm, although the humeri are relatively unaffected. The long bones of the lower extremity also exhibit bending, and the pelvis is flattened in the anteroposterior axis. All these features support a diagnosis of osteomalacia, but the specimen is that of a male, so the problem cannot be associated with nutritional deficiencies that can occur during childbearing. An additional complicating feature is the extensive development of abnormal bone on both femora and in the interosseous areas of the radius/ulna and tibia/fibula. This is not typical of osteomalacia and probably represents a pathological complication in addition to vitamin D deficiency.

Cases of rickets have been reported at several

archaeological sites in Europe for the Mesolithic period (Zivanovic 1975: 174; Nemeskéri and Lengyel 1978: 241; Grimm 1984; Meiklejohn and Zvelebil 1991), and the later Bronze Age (Schultz 1990: 178, 1993; Schultz and Schmidt-Schultz 1994). Reports of possible cases have also been recorded in the Middle East as early as the Mesolithic period (Macchiarelli 1989: 587). There may be additional cases in the Neolithic period (Röhler-Ertl 1981, as cited in Smith, Bar-Yosef, and Sillen 1984: 121) and at two sites in Dynastic Egypt (Ortner and Putschar 1981: 285; Buikstra, Baker, and Cook 1993: 44–5). In South Asia, there have been reports of rickets from the Mesolithic, Chalcolithic, and Iron Age periods (Lovell and Kennedy 1989: 91). Osteomalacia has been reported for Mesolithic sites in Europe (Nemeskéri and Lengyel 1978: 241) and in the Middle East (Macchiarelli 1989: 587).

Vitamin C Deficiency

Vitamin C (ascorbic acid) deficiency causes scurvy, a condition that is seen in both children and adults. Because humans cannot store vitamin C in the body, regular intake is essential. As vitamin C is abundant in fresh fruits and vegetables and occurs in small quantities in uncooked meat, scurvy is unlikely to occur in societies where such foods are common in the diet year-round. Historically, vitamin C deficiency has been endemic in northern and temperate climates toward the end of winter (Maat 1986: 160). In adults, scurvy is expressed only after four or five months of total deprivation of vitamin C (Stuart-Macadam 1989b: 219–20).

Vitamin C is critical in the formation of connective tissue, including bone protein and the structural proteins of blood vessels. In bone, the lack of vitamin C may lead to diminished bone protein (osteoid) formation by osteoblasts. The failure to form osteoid results in the abnormal retention of calcified cartilage, which has less biomechanical strength than normal bone. Fractures, particularly at the growth plate, are a common feature. In blood vessel formation the vessel walls may be weak, particularly in young children. This defect may result in bleeding from even minimal trauma. Bleeding can elevate the periosteum and lead to the formation of abnormal subperiosteal bone. It can also stimulate an inflammatory response resulting in abnormal bone destruction or formation adjacent to the bleeding.

Reports of scurvy in archaeological human remains are not common for several reasons. First, evidence of scurvy is hard to detect. For example, if scurvy is manifested in the long bones of a population, the frequency will probably represent only half of the actual cases (Maat 1986: 159). Second, many of the anatomical features associated with scurvy are as yet poorly understood, as is illustrated by an unusual type and distribution pattern of lesions being studied by Ortner. The pattern occurs in both Old and New World specimens in a variety of stages of severity

(e.g., Ortner 1984). Essentially, the lesions are inflammatory and exhibit an initial stage that tends to be destructive, with fine porous holes penetrating the outer table of the skull. In later stages the lesions are proliferative but tend to be porous and resemble lesions seen in the anemias. However, the major distinction from the anemias is that the diploë is not involved in the scorbutic lesions and the anatomical distribution in the skull tends to be limited to those areas that lie beneath the major muscles associated with chewing – the temporalis and masseter muscles.

An interesting Old World case of probable scurvy is from the cemetery for the medieval hospital of St. James and St. Mary Magdalene in Chichester, England. Throughout much of the medieval period the hospital was for lepers. As leprosy declined in prevalence toward the end of the period, patients with other ailments were admitted.

The specimen (Chichester burial 215) consists of the partial skeleton of a child about 6 years old, probably from the latter part of the medieval period. The only evidence of skeletal pathology occurs in the skull, where there are two types of lesion. The first type is one in which fine holes penetrate the compact bone with no more than minimal reactive bone formation. This condition is well demonstrated in bone surrounding the infraorbital foramen, which provides a passageway for the infraorbital nerve, artery, and vein. In the Chichester child there are fine holes penetrating the cortical bone on the margin of the foramen (Figure I.2.1) with minimal reactive bone formation. The lesion is indicative of chronic inflammation that could have been caused by blood passing through the walls of defective blood vessels.



Figure I.2.1. External view of the maxilla of a child about 6 years of age at the time of death. The right infra-orbital foramen exhibits an area of porosity (arrow) with slight evidence of reactive bone formation. Below the foramen is an area of post-mortem bone loss that is unrelated to the antemortem porosity. This burial (no. 215) is from the cemetery of the medieval Hospital of St. James and St. Mary Magdalene in Chichester, England.

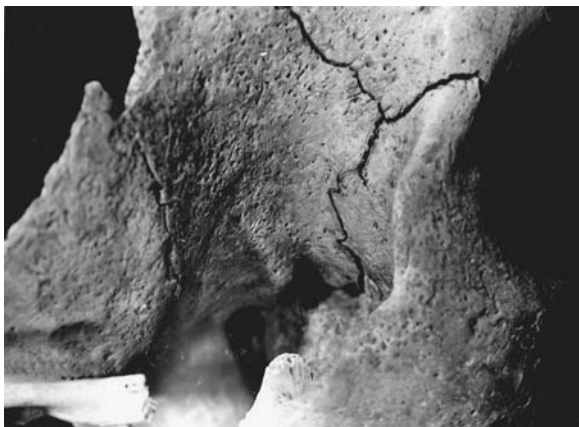


Figure I.2.2. Right sphenoid and adjacent bone surfaces of case seen in Figure I.2.1. Note the abnormal porosity of cortical bone.



Figure I.2.3. Orbital roof of case seen in Figure I.2.1. Note the enlarged, irregular, and porous surface.

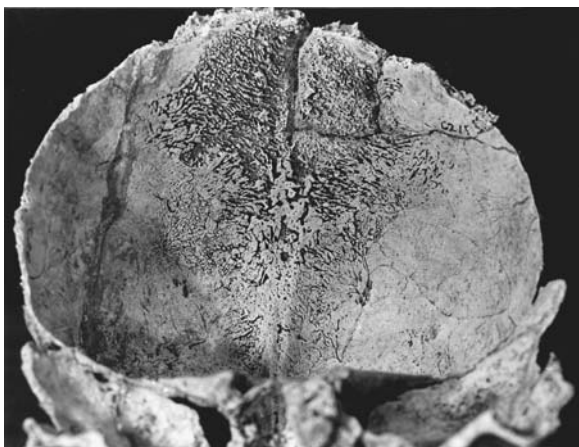


Figure I.2.4. Inner table of the frontal bone of case seen in Figure I.2.1. Note the large area of porous new bone formation.

Another area of porosity is apparent bilaterally on the greater wing of the sphenoid and adjacent bone tissue (Figure I.2.2). This area of porosity underlies the temporalis muscle, which has an unusual vascular supply that is particularly vulnerable to mild trauma and bleeding from defective blood vessels.

The second type of lesion is characterized by porous, proliferative lesions and occurs in two areas. One of these areas is the orbital roof (Figure I.2.3). At this site, bilateral lesions, which are superficial to the normal cortex, are apparent. The surfaces of the pathological bone tissue, particularly in the left orbit, seem to be filling in the porosity, suggesting that recovery from the pathological problem was in progress at the time of death.

The second area of abnormal bone tissue is the internal cortical surface of the skull, with a particular focus in the regions of the sagittal and transverse venous sinuses (Figure I.2.4). Inflammation, perhaps due to chronic bleeding between the dura and the inner table because of trauma to weakened blood vessels, is one possible explanation for this second type of lesion, particularly in the context of lesions apparent in other areas of the skull.

The probable diagnosis for this case is scurvy, which is manifested as a bone reaction to chronic bleeding from defective blood vessels. This diagnosis is particularly likely in view of the anatomical location of the lesions, although there is no evidence of defective bone tissue in the growth plates of the long bones (trümmerfeld zone) as one would expect in active scurvy. However, this may be the result of partial recovery from the disease as indicated by the remodeling in the abnormal bone tissue formed on the orbital roof.

The Chichester case provides probable evidence of scurvy in medieval England. C. A. Roberts (1987) has reported a case of possible scurvy from a late Iron Age or early Roman (100 B.C. to A.D. 43) site in Beckford, Worcestershire, England. She described an infant exhibiting porous proliferative orbital lesions and reactive periostitis of the long bones. Schultz (1990: 178) has discussed the presence of infantile scurvy in Bronze Age sites in Europe (2200 to 1900 B.C.) and in Anatolia (2500 to 2300 B.C.) (Schultz and Schmidt-Schultz 1994: 8). In South Asia pathological cases possibly attributable to infantile scurvy have been reported in Late Chalcolithic/Iron Age material (Lukacs and Walimbe 1984: 123).

Iron Deficiency

Iron deficiency is today a common nutritional problem in many parts of the world. Two-thirds of women and children in developing countries are iron deficient (Scrimshaw 1991: 46). However, physical evidence for this condition in antiquity remains elusive, and detection of trends in space and time remain inconclusive.

There are two general types of anemia that affect the human skeleton. Genetic anemias, such as sickle cell anemia and thalassemia, are caused by defects in red blood cells. Acquired anemias may result from chronic bleeding (such as is caused by internal parasites), or from an infection that will lead to a state of anemia (Stuart-Macadam 1989a; Meiklejohn and Zvelebil 1991: 130), or from an iron-deficient diet. Deficient dietary iron can be the result of either inadequate intake of iron from dietary sources or failure to absorb iron during the digestion of food.

Iron is a critical element in hemoglobin and important in the transfer and storage of oxygen in the red blood cells. Defective formation of hemoglobin may result in an increased turnover of red blood cells; this greatly increases demand for blood-forming marrow. In infants and small children the space available for blood formation is barely adequate for the hematopoietic marrow needed for normal blood formation. Enlargement of hematopoietic marrow space can occur in any of the bones. In long bones, marrow may enlarge at the expense of cortical bone, creating greater marrow volume and thinner cortices. In the skull, anemia produces enlargement of the diploë, which may replace the outer table, creating very porous bone tissue known as porotic hyperostosis. Porotic hyperostosis is a descriptive term first used by J. L. Angel in his research on human remains in the eastern Mediterranean (1966), where it is a well-known condition in archaeological skeletal material. Porotic enlargement of the orbital roof is a specific form of porotic hyperostosis called *cribra orbitalia*. The presence of both these conditions has been used by paleopathologists to diagnose anemias in archaeological human remains.

Attributing porotic hyperostosis to anemia should be done with caution for several reasons. First, diseases other than anemia (i.e., scurvy, parasitic infection, and rickets) can cause porotic enlargement of the skull. There are differences in pathogenesis that cause somewhat different skeletal manifestations, but overlap in pathological anatomy is considerable. Second, as mentioned previously, some diseases such as scurvy may occur in addition to anemia. Because both diseases cause porotic, hypertrophic lesions of the skull, careful anatomical analysis is critical. Finally, attributing porotic hyperostosis to a specific anemia, such as iron-deficiency anemia, is problematic. On the basis of anatomical features alone, it is very difficult to distinguish the bone changes caused by lack of iron in the diet from bone changes caused by one of the genetic anemias. These cautionary notes are intended to highlight the need for care in interpreting published reports of anemia (and other diseases caused by malnutrition), particularly when a diagnosis of a specific anemia is offered.

Angel (1966, 1972, 1977) was one of the earliest observers to link porotic hyperostosis in archaeologi-

cal human remains to genetic anemia (thalassemia). He argued that thalassemia was an adaptive mechanism in response to endemic malaria in the eastern Mediterranean. The abnormal hemoglobin of thalassemia, in inhibiting the reproduction of the malarial parasite, protects the individual from severe disease.

As indicated earlier, in malarial regions of the Old World, such as the eastern Mediterranean, it may be difficult to differentiate porotic hyperostosis caused by genetic anemia from dietary anemia. However, in nonmalarial areas of the Old World, such as northern Europe, this condition is more likely to be caused by nongenetic anemia such as iron-deficiency anemia.

Because determining the probable cause of anemia is so complex, few reports have been able to provide a link between porotic hyperostosis and diet. In prehistoric Nubian populations, poor diet may have been one of the factors that led to iron-deficiency anemia (Carlson et al. 1974, as cited in Stuart-Macadam 1989a: 219). At Bronze Age Toppo Daguzzo in Italy (Repetto, Canci, and Borgogni Tarli 1988: 176), the high rate of *cribra orbitalia* was, possibly, caused by nutritional stress connected with weaning. At Metaponto, a Greek colony (c. 600 to 250 B.C.) in southern Italy noted for its agricultural wealth, the presence of porotic hyperostosis, along with other skeletal stress markers, indicated to researchers that the colony had nutritional problems (Henneberg, Henneberg, and Carter 1992: 452). It has been suggested that specific nutrients may have been lacking in the diet.

Fluorosis

Fluorosis as a pathological condition occurs in geographical regions where excessive fluorine is found in the water supply. It may also occur in hot climates where spring or well water is only marginally high in fluoride, but people tend to drink large amounts of water, thereby increasing their intake of fluoride. In addition, high rates of evaporation may increase the concentration of fluoride in water that has been standing (Littleton and Frohlich 1993: 443). Fluorosis has also been known to occur where water that contains the mineral is used to irrigate crops or to prepare food, thereby increasing the amount ingested (Leverett 1982, as cited in Lukacs, Retief, and Jarrige 1985: 187).

In the Old World, fluorosis has been documented in ancient populations of Hungary (Molnar and Molnar 1985: 55), India (Lukacs et al. 1985: 187), and areas of the Arabian Gulf (Littleton and Frohlich 1993: 443).

Fluorosis is known primarily from abnormalities of the permanent teeth, although the skeleton may also be affected. If excessive fluorine is ingested during dental development, dentition will be affected in several ways, depending upon severity. J. Littleton and B. Frohlich (1989: 64) observed fluorosis in archaeological specimens from Middle Bronze Age

and Islamic periods in Bahrain. They categorized their findings into four stages of severity: (1) normal or translucent enamel, (2) white opacities on the enamel, (3) minute pitting with brownish staining, and (4) finally, more severe and marked pitting with widespread brown to black staining of the tooth. They noted that about 50 percent of the individuals in both the Bronze Age and the Islamic periods showed dental fluorosis (1989: 68).

Other cases of dental fluorosis have been reported in the archaeological record. At a site in the Arabian Gulf on the island of Umm an Nar (c. 2500 B.C.), Littleton and Frohlich (1993: 443) found that 21 percent of the teeth excavated showed signs of fluorosis. In Hungary, S. Molnar and I. Molnar (1985) reported that in seven skeletal populations dated from late Neolithic to late Bronze Age (c. 3000 B.C. to c. 1200 B.C.), "mottled" or "chalky" teeth suggestive of fluorosis appeared. The frequencies varied from 30 to 67 percent of individuals (Molnar and Molnar 1985: 60). In South Asia, J. R. Lukacs, D. H. Retief, and J. F. Jarringe (1985: 187) found dental fluorosis at Early Neolithic (c. 7000 to 6000 B.C.) and Chalcolithic (c. 4000 to 3000 B.C.) levels at Mehgarh.

In order for fluoride to affect the skeletal system, the condition must be long-standing and severe (Flemming Møller and Gudjonsson 1932; Sankaran and Gadekar 1964). Skeletal manifestations of fluorosis may involve ossification of ligament and tendon tissue at their origin and insertion (Figure I.2.5). However, other types of connective tissue may also be ossified, such as the tissue at the costal margin of the ribs. Connective tissue within the neural canal is involved in some cases, reducing the space needed for the spinal cord and other neurological pathways. If severe, it can cause nerve damage and paralysis.

Fluorosis may also affect mineralization of osteoid during osteon remodeling in the microscopic structure of bone (Figure I.2.6a). In contrast with the ossification of ligament and tendon tissue, excessive fluorine inhibits mineralization of osteoid at the histological level of tissue organization. It is unclear why, in some situations, fluorosis stimulates abnormal mineralization, yet in other situations, it inhibits mineralization. In microradiographs, inhibited mineralization is seen as a zone of poor mineralization (Figure I.2.6b).

Examples of archaeological fluorosis of bone tissue are rare. However, in Bahrain, excavations from third to second millennium B.C. burial mounds have revealed the full range of this disease (Frohlich, Ortner, and Al-Khalifa 1987/88). In addition to dental problems, skeletons show ossification of ligaments and tendons, and some exhibit ossification of connective tissue within the neural canal. The most severe case is that of a 50-year-old male who had a fused spine in addition to large, bony projections in the ligament attachments of the radius and ulna and tibia and fibula. Chemical analysis indicates almost 10 times the normal levels of fluorine in bone material.



Figure I.2.5. Right lateral view of the ninth through the twelfth thoracic vertebrae from the skeleton of a male about 45 years of age at the time of death. Note the fusion of the vertebral bodies and the ossification of the interspinous ligaments (arrow). Burial (B South 40) is from the Early Bronze Age (c. 2100 B.C.) site of Madinat Hamad in Bahrain.

Protein–Energy Malnutrition

Protein–energy malnutrition (PEM), or protein–calorie deficiency, covers a range of syndromes from malnutrition to starvation. The best-known clinical manifestations are seen in children in the form of kwashiorkor (a chronic form that is caused by lack of protein) and marasmus (an acute form where the child wastes away) (Newman 1993: 950). PEM occurs in areas of poverty, with its highest rates in parts of Asia and Africa (Newman 1993: 954).

PEM has no specific skeletal markers that enable us to identify it in skeletal remains. It affects the human skeleton in different ways, depending on severity and age of occurrence. During growth and development it may affect the size of the individual so that bones and teeth are smaller than normal for that population. There may be other manifestations of PEM during this growth period, such as diminished sexual dimorphism, decreased cortical bone thickness, premature osteoporosis (associated with starvation),

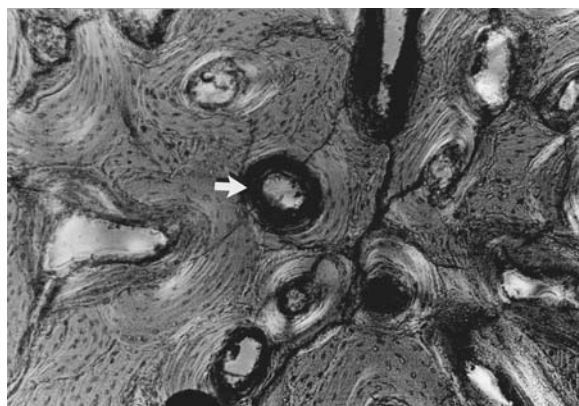


Figure I.2.6a. Photomicrograph of a bone section from the femur of the burial seen in Figure I.2.5. The dark zone (arrow) in several osteons is an area of poor mineralization caused by the toxic effects of excessive fluorine, approximately 125× before reduction.

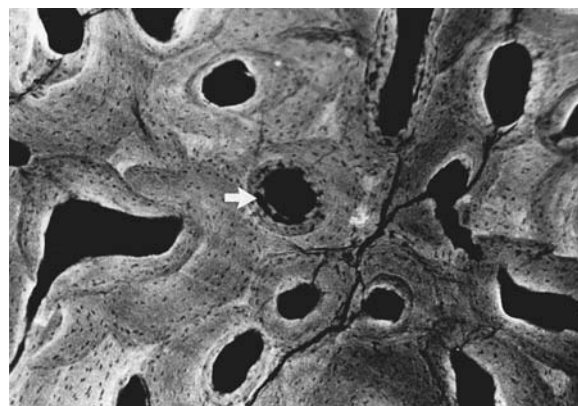


Figure I.2.6b. Photomicrograph of a microradiograph of the bone section seen in Figure I.2.6a. Note the areas of low density (dark to black) that correspond to the dark areas seen in Figure I.2.6a. The arrow indicates one of these areas.

enamel hypoplasias, and Harris lines. Because malnutrition decreases the immune response to infection, a high rate of infection may also indicate nutritional problems. Unfortunately, most of the indicators of growth problems in malnutrition occur in other disease syndromes as well; thus, careful analysis of subtle abnormalities in skeletal samples is needed. Chemical and histological analyses provide supporting evidence of abnormal features apparent anatomically.

PEM is probably as old as humankind (Newman 1993: 953). Written records in the Old World over the past 6,000 years have alluded to frequent famines. Beginning around 4000 B.C. and ending around 500 B.C., the Middle East and northeastern Africa, specifically the Nile and Tigris and Euphrates river valleys, were “extraordinarily famine prone” (Dirks 1993: 162). The skeletal evidence in archaeological remains is based on a number of skeletal abnormalities that, observers have concluded, are the result of nutritional problems.

Several studies suggesting problems with nutrition have been undertaken in northeastern Africa. In reviewing 25 years of work done on prehistoric Nubian skeletal material, G. J. Armelagos and J. O. Mills (1993: 10–11) noted that reduced long bone growth in children and premature bone loss in both children and young women were due to nutritional causes, specifically to impaired calcium metabolism. One of the complications of PEM in modern populations is thought to be interference with the metabolism of calcium (Newman 1993: 953). In Nubia, reliance on cereal grains such as barley, millet, and sorghum, which are poor sources of calcium and iron, may have been a major factor in the dietary deficiency of the population (Armelagos and Mills 1993: 11). In a later Meroitic site (c. 500 B.C. to A.D. 200) in the Sudan, E. Fulcheri and colleagues (1994: 51) found that 90 percent of the

children’s skeletons (0 to 12 years old) showed signs of growth disturbances or nutritional deficiencies.

There are signs of malnutrition from other areas and time periods as well. In the Arabian Gulf, the Mesolithic necropolis (c. 3700 to 3200 B.C.) of Ra’s al Hamra revealed skeletal remains of a population under “strong environmental stress” with numerous pathologies, including rickets, porotic hyperostosis, and cribra orbitalia (Macchiarelli 1989). In addition, indications of growth disturbances in the form of a high rate of enamel hypoplasias and a low rate of sexual dimorphism have led to the conclusion that part of this stress was nutritional (Coppa, Cucina, and Mack 1993: 79). Specifically, the inhabitants may have suffered from protein–calorie malnutrition (Macchiarelli 1989: 587).

At Bronze Age (third millennium B.C.) Jelsoyce in Slovakia, M. Schultz and T. H. Schmidt-Schultz (1994: 8) found “strong evidence of malnutrition” for the infant population, but noted that the relatively high frequency of enamel hypoplasia, anemia, rickets, and scurvy, in addition to infection, was not typical for the Bronze Age. Nutritional stress has also been suggested by the presence of premature osteoporosis among the pre-Hispanic inhabitants of the Canary Islands (Reimers et al. 1989; Martin and Mateos 1992) and among the population of Bronze Age Crete (McGeorge and Mavroudis 1987).

Conclusion

A review of the literature combined with our own research and experience leaves no doubt in our minds that humans have had nutritional problems extending at least back into the Mesolithic period. We have seen probable evidence of vitamin C deficiency, vitamin D deficiency, iron-deficiency anemia, fluorosis, and protein–energy malnutrition. However, because the

conditions that cause malnutrition may be sporadic or even random, they vary in expression in both time and space. The prevalence of nutritional diseases may be due to food availability that can be affected by local or seasonal environment. For example, crop failure can result from various factors, such as shortage of water or overabundance of pests. Other nutritional problems can be caused by idiosyncratic circumstances such as individual food preferences or specific cultural customs.

Culture affects nutrition in influencing the foods that are hunted, gathered, herded, or cultivated, as well as the ways they are prepared for consumption. Cultural traditions and taboos frequently dictate food choices. All these variables affecting nutrition, combined with differences in observers and the varying methodologies they use in studying ancient human remains, make finding diachronic patterns or trends in human nutrition difficult.

Whether or not humankind benefited from or was harmed by the epochal changes in the quality and quantity of food over the past 10,000 years is, in our opinion, still open to debate. Many studies of skeletal remains conclude that the level of health, as indicated by nutrition, declined with the change from the Mesolithic hunter-gatherer way of life to the later period of developed agriculture. M. N. Cohen and G. J. Armelagos (1984a: 587), in summing up the results of a symposium on the paleopathology of the consequences of agriculture, noted that studies of both the Old and New Worlds provided consistent evidence that farming was accompanied by a decline in the quality of nutrition. Other, more recent studies have indicated agreement with this conclusion. A. Agelarakis and B. Waddell (1994: 9), working in southwestern Asia, stated that skeletal remains from infants and children showed an increase in dietary stress during the agricultural transition. Similarly, N. C. Lovell and K. A. R. Kennedy (1989: 91) observed that signs of nutritional stress increased with farming in South Asia.

By contrast, however, in a thorough review of well-studied skeletal material from Mesolithic and Neolithic Europe, C. Meiklejohn and M. Zvelebil (1991) found unexpected variability in the health status of populations connected with the Mesolithic-Neolithic transition. Part of this variability was related to diet, and they concluded that for Europe, no significant trends in health were visible in the skeletons of those populations that made the transition from hunting and gathering to greater dependence on agriculture, and from mobile to relatively sedentary communities. Although some differences between specific areas (i.e., the western Mediterranean and northern and eastern Europe) seem to exist, deficiencies in sample size mean that neither time- nor space-dependent patterns emerge from their review of the existing data. Clearly, different observers interpret evidence on the history of nutritional diseases in somewhat different ways. This is not surprising given the nature of the data. The questions about the relationship of malnutrition to

changes in time and space remain an important scientific problem. Additional studies on skeletal material, particularly those that apply new biochemical and histological methods, offer the promise of a clearer understanding of these issues in the near future.

Donald J. Ortner
Gretchen Theobald

Research support for D. J. Ortner was provided by grants from the Smithsonian Scholarly Studies Program (fund no. 1233S40C). The assistance of Agnes Stix, in preparation of photographic illustrations, and editing, and Janet T. Beck for editing, Department of Anthropology, National Museum of Natural History, is deeply appreciated.

Bibliography

- Agelarakis, A., and B. Waddell. 1994. Analysis of non-specific indicators of stress in subadult skeletons during the agricultural transition from southwestern Asia. *Homo* 45 (Supplement): 9.
- Angel, J. L. 1966. Porotic hyperostosis, anemias, malarial and marshes in the prehistoric eastern Mediterranean. *Science* 153: 760-3.
1971. Early Neolithic skeletons from Catal Huyuk: Demography and pathology. *Anatolian Studies* 21: 77-98.
1972. Biological relations of Egyptian and eastern Mediterranean populations during pre-Dynastic and Dynastic times. *Journal of Human Evolution* 1: 307-13.
1977. Anemias of antiquity in the eastern Mediterranean. *Paleopathology Association Monograph No. 2*: 1-5.
- Armelagos, George J., and James O. Mills. 1993. Paleopathology as science: The contribution of Egyptology. In *Biological anthropology and the study of ancient Egyptians*, ed. W. Vivian Davies and Roxie Walker, 1-18. London.
- Bennike, P. 1985. *Palaeopathology of Danish skeletons*. Copenhagen.
- Brothwell, Don. 1986. *The Bog Man and the archaeology of people*. London.
- Buikstra, Jane E., Brenda J. Baker, and Della C. Cook. 1993. What diseases plagued ancient Egyptians? A century of controversy considered. In *Biological anthropology and the study of ancient Egypt*, ed. W. Vivian Davies and Roxie Walker, 24-53. London.
- Carli-Thiele, P., and Michael Schultz. 1994. Cribra orbitalia in the early neolithic child populations from Wandersleben and Aiterhofen (Germany). *Homo* 45 (Supplement): 33.
- Cohen, Mark Nathan. 1989. *Health and the rise of civilization*. New Haven, Conn.
- Cohen, Mark Nathan, and George J. Armelagos. 1984a. Paleopathology at the origins of agriculture: Editors' summation. In *Paleopathology at the origins of agriculture*, ed. Mark Nathan Cohen and George J. Armelagos, 585-601. New York.
- eds. 1984b. *Paleopathology at the origins of agriculture*. New York.
- Coppa, A., A. Cucina, and M. Mack. 1993. Frequenza e distribuzione cronologica dell' ipoplasia dello smalto in un campione scheletrico di RH5. *Antropologia Contemporanea* 16: 75-80.
- Dirks, Robert. 1993. Famine and disease. In *The Cambridge world history of human disease*, ed. K. F. Kiple, 157-63. Cambridge and New York.
- Fallon, M. D. 1988. Bone histomorphology. In *Diagnosis of*

- bone and joint disorders. Second edition, ed. D. Resnick and G. Niwayama, 1975-97. Philadelphia, Pa.
- Fischer, Christian. 1980. Bog bodies of Denmark. In *Mummies, disease, and ancient cultures*, ed. Aidan Cockburn and Eve Cockburn, 177-93. New York.
- Flemming Møller, P., and S. V. Gudjonsson. 1932. Massive fluorosis of bones and ligaments. *Acta Radiologica* 13: 269-94.
- Follis, R. H., Jr., D. A. Jackson, and E. A. Park. 1940. The problem of the association of rickets and scurvy. *American Journal of Diseases of Children* 60: 745-7.
- Frohlich, Bruno, Donald J. Ortner, and Haya Ali Al-Khalifa. 1987/1988. Human disease in the ancient Middle East. *Dilmun: Journal of the Bahrain Historical and Archaeological Society* 14: 61-73.
- Fulcheri, E., P. Baracchini, A. Coppa, et al. 1994. Paleopathological findings in an infant Meroitic population of El Geili (Sudan). *Homo* 45 (Supplement): 51.
- Glob, P. V. 1971. *The bog people*. London.
- Goldberg, A. 1963. The anaemia of scurvy. *Quarterly Journal of Medicine* 31: 51-64.
- Goodman, A. H. 1991. Stress, adaptation, and enamel developmental defects. In *Human paleopathology: Current syntheses and future options*, ed. D. J. Ortner and A. C. Aufderheide, 280-7. Washington, D.C.
- Goodman, A. H., D. L. Martin, and G. J. Armelagos. 1984. Indications of stress from bone and teeth. In *Paleopathology at the origins of agriculture*, ed. Mark Nathan Cohen and George J. Armelagos, 13-49. Orlando, Fla.
- Grimm, H. 1984. Neue Hinweise auf ur- und frühgeschichtliches sowie mittelalterliches Vorkommen der Rachitis und ähnlicher Mineralisationsstörungen. *Ärztliche Jugendkunde* 5: 168-77.
- Grmek, M. D. 1989. *Diseases in the ancient Greek world*, trans. Mirelle Muellner and Leonard Muellner. Baltimore, Md.
- Hall, Thomas L., and Victor W. Sidel. 1993. Diseases of the modern period in China. In *The Cambridge world history of human disease*, ed. K. F. Kiple, 362-73. Cambridge and New York.
- Henneberg, M., R. Henneberg, and J. C. Carter. 1992. Health in colonial Metaponto. *Research and Exploration* 8: 446-59.
- Katzenberg, M. Anne. 1992. Advances in stable isotope analysis of prehistoric bones. In *Skeletal biology of past peoples: Research methods*, ed. Shelly R. Saunders and M. Anne Katzenberg, 105-19. New York.
- Kuhnke, LaVerne. 1993. Disease ecologies of the Middle East and North Africa. In *The Cambridge world history of human disease*, ed. K. F. Kiple, 453-62. Cambridge and New York.
- Littleton, J., and B. Frohlich. 1989. An analysis of dental pathology and diet on historic Bahrain. *Paleorient* 15: 59-75.
1993. Fish-eaters and farmers: Dental pathology in the Arabian Gulf. *American Journal of Physical Anthropology* 92: 427-47.
- Lovell, N. C., and K. A. R. Kennedy. 1989. Society and disease in prehistoric South Asia. In *Old problems and new perspectives in the archaeology of South Asia*, ed. M. Kenoyer, 89-92. Madison, Wis.
- Lukacs, John R. 1989. Dental paleopathology: Methods of reconstructing dietary patterns. In *Reconstruction of life from the skeleton*, ed. M. Y. Iscan and K. A. R. Kennedy, 261-86. New York.
- Lukacs, John R., D. H. Retief, and J. F. Jarrige. 1985. Dental disease in prehistoric Baluchistan. *National Geographic Research* 1: 184-97.
- Lukacs, John R., and Subhash R. Walimbe. 1984. Paleodemography at Inamgaon: An early farming village in western India. In *The people of South Asia: The biological anthropology of India, Pakistan, and Nepal*, ed. John R. Lukacs, 105-33. New York.
- Maat, George J. R. 1986. Features of malnutrition, their significance and epidemiology in prehistoric anthropology. In *Innovative trends in prehistoric anthropology*, ed. B. Hermann, 157-64. Berlin.
- Macchiarelli, R. 1989. Prehistoric "fisheaters" along the eastern Arabian coasts: Dental variation, morphology, and oral health in the Ra's al-Hamra community (Qurum, Sultanate of Oman, 5th-4th millennia B.C.). *American Journal of Physical Anthropology* 78: 575-94.
- Martin, C. Rodriguez, and B. Beranger Mateos. 1992. Interpretation of the skeletal remains from Los Auchones (Anaga, Santa Cruz de Tenerife): A case of biocultural isolation. *Papers on paleopathology presented at the 9th European members meeting*: 22-3. Barcelona.
- McGeorge, P. J. P., and E. Mavroudis. 1987. The incidence of osteoporosis in Bronze Age Crete. *Journal of Paleopathology* 1: 37.
- Meiklejohn, Christopher, and Marek Zvelebil. 1991. Health status of European populations at the agricultural transition and the implications for the adoption of farming. In *Health in past societies: Biocultural interpretations of human skeletal remains in archaeological contexts*, ed. H. Bush and M. Zvelebil, 129-45. Oxford.
- Molnar, S., and I. Molnar. 1985. Observations of dental diseases among prehistoric populations in Hungary. *American Journal of Physical Anthropology* 67: 51-63.
- Nemeskéri, J., and I. Lengyel. 1978. The results of paleopathological examinations. In *Vlasac: A Mesolithic settlement in the Iron Gates*, Vol. 2, ed. M. Garasanin, 231-60. Belgrade.
- Newman, James L. 1993. Protein-energy malnutrition. In *The Cambridge world history of human disease*, ed. K. F. Kiple, 950-5. Cambridge and New York.
- Ortner, D. J. 1984. Bone lesions in a probable case of scurvy from Metlatavik, Alaska. *MASCA* 3: 79-81.
- Ortner, D. J., and W. G. J. Putschar. 1981. *Identification of pathological conditions in human skeletal remains*. Washington, D.C.
- Reimers, C., Emilio Gonzales, Matilde Arnay De La Rosa, et al. 1989. Bone histology of the prehistoric inhabitants of Gran Canaria. *Journal of Paleopathology* 2: 47-59.
- Repetto, E., A. Canci, and S. M. Borgogni Tarli. 1988. Skeletal indicators of health conditions in the Bronze Age sample from Toppo Daguzzo (Basilicata, Southern Italy). *Anthropologie* 26: 173-82.
- Resnick, D., and G. Niwayama. 1988. *Diagnosis of bone and joint disorders*. Second edition. Philadelphia, Pa.
- Roberts, C. A. 1987. Case report no. 9. *Paleopathology Newsletter* 57: 14-15.
- Rowling, J. T. 1967. Paraplegia. In *Diseases in antiquity: A survey of the diseases, injuries and surgery of early populations*, ed. D. R. Brothwell and A. T. Sandison, 272-8. Springfield, Ill.
- Sankaran, B., and N. Gadekar. 1964. Skeletal fluorosis. In *Proceedings of the first European bone and tooth symposium*, Oxford, 1963. Oxford.
- Schultz, Michael. 1986. *Die mikroskopische Untersuchung prähistorischer Skelettfunde*. Liestal, Switzerland.
1990. Causes and frequency of diseases during early childhood in Bronze Age populations. In *Advances in paleopathology*, ed. L. Capasso, 175-9. Chieti, Italy.
1993. Initial stages of systemic bone disease. In *Histology*

- of ancient bone, ed. G. Grupe and A. N. Garland, 185–203. Berlin.
- Schultz, Michael, and T. H. Schmidt-Schultz. 1994. Evidence of malnutrition and infectious diseases in the infant population of the early Bronze Age population from Jelsovce (Slovakia): A contribution to etiology and epidemiology in prehistoric populations. *Papers on paleopathology presented at the 21st Annual Meeting of the Paleopathology Association*: 8. Denver.
- Scrimshaw, Nevin S. 1991. Iron deficiency. *Scientific American* 265: 46–52.
- Smith, P., O. Bar-Yosef, and A. Sillen. 1984. Archaeological and skeletal evidence for dietary change during the Late Pleistocene/Early Holocene in the Levant. In *Paleopathology at the origins of agriculture*, ed. Mark Nathan Cohen and George J. Armelagos, 101–36. Orlando, Fla.
- Stuart-Macadam, P. 1989a. Nutritional deficiency diseases: A survey of scurvy, rickets, and iron-deficiency anemia. In *Reconstruction of life from the skeleton*, ed. M. Y. Iscan and K. A. R. Kennedy, 201–22. New York.
- 1989b. Rickets as an interpretive tool. *Journal of Paleopathology* 2: 33–42.
- White, Christine D. 1993. Isotopic determination of seasonality in diet and death from Nubian mummy hair. *Journal of Archaeological Science* 20: 657–66.
- Zivanovic, S. 1975. A note on the anthropological characteristics of the Padina population. *Zeitschrift für Morphologie und Anthropologie* 66: 161–75.

1.3. Dietary Reconstruction As Seen in Coprolites

The question of prehistoric dietary practices has become an important one. Coprolites (desiccated or mineralized feces) are a unique resource for analyzing prehistoric diet because their constituents are mainly the undigested or incompletely digested remains of food items that were actually eaten. Thus they contain direct evidence of dietary intake (Bryant 1974b, 1990; Spaulding 1974; Fry 1985; Scott 1987; Sobolik 1991a, 1994a, 1994b). In addition they can reveal important information on the health, nutrition, possible food preparation methods, and overall food economy and subsistence of a group of people (Sobolik 1991b; Reinhard and Bryant 1992).

Coprolites are mainly preserved in dry, arid environments or in the frozen arctic (Carbone and Keel 1985). Caves and enclosed areas are the best places for preserved samples and there are also samples associated with mummies. Unfortunately, conditions that help provide such samples are not observed in all archaeological sites.

Coprolite analysis is important in the determination of prehistoric diets for two significant reasons. First, the constituents of a coprolite are mainly the remains of intentionally eaten food items. This type of precise sample cannot be replicated as accurately from animal or plant debris recovered from archaeological sites. Second, coprolites tend to preserve

small, fragile remains, mainly because of their compact nature, which tends to keep the constituents separated from the site matrix. These remains are typically recovered by normal coprolitic processing techniques, which involve screening with micron mesh screens rather than the larger screens used during archaeological excavations.

The limitations of coprolites are also twofold (Sobolik 1994a). First, even though the analysis of coprolites indicates the ingestion of food items, their constituents do not contain the entire diet of an individual or a population. In fact, because the different food items ingested pass through the digestive system at different rates, coprolite contents do not reflect one specific meal. The problem with coprolites is that they contain the indigestible portion of foods. The actual digested portion has been absorbed by the body. Thus, it has been estimated that meat protein may be completely absorbed during the digestion process, often leaving few traces in the coprolite (Fry 1985). However, recent protein residue analyses conducted on coprolites have indicated that some protein may survive (Newman et al. 1993).

A second limitation is that coprolites can reflect seasonal or short-term dietary intake. Individual coprolites often reflect either items a person ate earlier that day or what may have been eaten up to a month before (Williams-Dean 1978; Sobolik 1988). Thus, determining year-round dietary intake using coprolites, even with a large sample, becomes risky and inconclusive.

The History of Coprolite Research

The first observations of coprolites were those from animals of early geologic age; the Cretaceous in England (Mantell 1822; Agassiz 1833–43) and North America (Dekay 1830); the Lower Jurassic in England (Buckland 1829); and the Eocene in France (Robert 1832–33). Later works in North America include coprolites of the ground sloth (Laudermilk and Munz 1934, 1938; Martin, Sabels, and Shutler 1961; Thompson et al. 1980) and other Pleistocene animals (Davis et al. 1984).

The potential of human coprolites as dietary indicators was realized by J. W. Harshberger in 1896. The first analyses, however, were not conducted until after the beginning of the twentieth century. These initial studies were conducted by G. E. Smith and F. W. Jones (1910), who examined the dried fecal remains from Nubian mummies, and by B. H. Young (1910), L. L. Loud and M. R. Harrington (1929), and Volney H. Jones (1936), who studied materials from North American caves. Early coprolite analyses also included samples from Danger Cave (Jennings 1957), sites in Tamaulipas, Mexico (MacNeish 1958), caves in eastern Kentucky (Webb and Baby 1957), and colon contents from a mummy (Wakefield and Dellinger 1936).

The processing techniques for these early analyses

consisted of either cutting open the dry coprolites and observing the large, visible contents, or grinding the samples through screens, in the process breaking much of the material. Improved techniques for analyzing coprolites were later developed by Eric O. Callen and T.W.M. Cameron (1960).

Still used today, these techniques revolutionized the science of coprolite analysis. They involved rehydrating the sample in tri-sodium phosphate, a strong detergent, in order to gently break apart the materials for ease in screening. Processing with tri-sodium phosphate also allowed for the recovery of polleniferous and parasitic materials from the samples, and increased the recovery of smaller, fragile macromaterials. Direct coprolite pollen analyses were soon followed by the first published investigation conducted by Paul S. Martin and F.W. Sharrock (1964) on material from Glen Canyon. Subsequently, there have been other innovative studies (Hill and Hevly 1968; Bryant 1974a; Bryant and Williams-Dean 1975; Hevly et al. 1979).

Coprolite Constituents

Coprolites represent such an unusual data source that their analysis is usually undertaken by specialists, generally by paleoethnobotanists (Ford 1979). A thorough coprolite analysis, however, involves the identification and interpretation of all types of botanical remains (Bryant 1974b, 1986; Fry 1985), such as fiber, seeds, and pollen, as well as nonbotanical remains, such as animal bone and hair, insects, fish and reptile scales, and parasites (to name a few of the many coprolitic constituents). Some recent studies have also identified the presence of wax and lipids through gas chromatography and mass spectrometry analyses and the analysis of phytoliths (Wales, Evans, and Leeds 1991; Danielson 1993). Clearly, then, coprolite analysis covers a myriad of sciences besides paleoethnobotany.

Yet, because coprolite analyses have tended to be conducted by paleoethnobotanists, such studies have tended to focus on the botanical remains. Recently, however, researchers have realized that the botanical portion represents a biased sample of prehistoric diet, and, consequently, studies of the nonbotanical macroremains from coprolites are becoming more prevalent.

A significant early analysis that included the identification and interpretation of a variety of coprolitic constituents was conducted on 50 coprolite samples from Lovelock Cave, Nevada (Napton 1970). As a part of this investigation, Charles L. Douglas (1969) identified eight different animal species through analysis of hair content, and Lewis K. Napton and O.A. Brunetti (1969) identified feathers from a wide variety of birds, most significantly, the mud hen.

A more recent study has focused on small animal remains recovered from coprolites excavated throughout North America (Sobolik 1993). This effort indicates that animals that have been considered noncultural or

site-contaminants actually served as human food (Munson, Parmalee, and Yarnell 1971; Parmalee, Paloumpis, and Wilson 1972; Smith 1975; Cordell 1977; Lyman 1982). The large number of coprolites analyzed from North America reveals direct ingestion of small animals, suggesting that small animal remains from sites do, in fact, reflect human dietary patterns, and that reptiles, birds, bats, and a large variety of rodents were an important and prevalent component of the prehistoric diet.

A variety of microremains can be analyzed from coprolites. These constituents include spores and fungi (Reinhard et al. 1989), bacteria (Stiger 1977), viruses (Williams-Dean 1978), and, recently, phytoliths (Bryant 1969; Cummings 1989). The most frequently analyzed microremains from coprolites, however, are pollen and parasites.

Pollen

Pollen is a unique resource in the analysis of coprolites because it can provide information not obtained from the macroremains. If a flower type is frequently ingested, the soft flower parts will most likely be digested. Pollen, depending on size and structure, becomes caught in the intestinal lumen, permitting it to be excreted in fecal samples for up to one month after ingestion. Therefore, the pollen content of coprolites does not reflect one meal, but can reflect numerous meals with a variety of pollen types (Williams-Dean 1978; Sobolik 1988).

Pollen in coprolites can occur through the intentional eating of flowers or seeds, through the unintentional ingestion of pollen in medicinal teas, or by the consumption of plants to which pollen adheres. Pollen, in this context, is considered "economic" because it is actually associated with food or a medicinal item. But it may also become ingested during respiration, with contaminated water supplies, and with food, especially if the food is prepared in an open area (Bryant 1974b, 1987). Such occurrences can be especially prevalent during the pollinating season of a specific plant, such as pine and oak in the spring or ragweed and juniper in the fall. Pollen, in this context, is considered "background" because it was accidentally ingested and was not associated with a particular food or medicinal item.

Pollen types are divided into insect pollinated plants (zoophilous) and wind pollinated plants (anemophilous). Insect pollinated plants produce few pollen grains and are usually insect specific to ensure a high rate of pollination. Indeed, such plants generally produce fewer than 10,000 pollen grains per anther (Faegri and Iversen 1964) and are rarely observed in the pollen record.

Wind pollinated plants, by contrast, produce large amounts of pollen to ensure pollination and are frequently found in the pollen record. The enormous quantity of pollen produced by some plants was highlighted by the study of R. N. Mack and Vaughn M.

Bryant (1974) in which they found over 50 percent *Pinus* pollen in areas where the nearest pine tree is more than 100 miles away. Knut Faegri and J. Iversen (1964) state that an average pine can produce approximately 350 million pollen grains per tree.

In coprolite analyses, this division between pollination types is essential because a high frequency of wind pollinated pollen types in a sample may indicate not diet but rather accidental ingestion from contaminated food or water supplies. A high frequency of insect pollinated pollen types, however, often indicates the intentional ingestion of food containing pollen (economic pollen), since it is unlikely that many grains of this type are accidental contaminants. Bryant (1975) has shown from field experiments that for some of the common insect pollinated types in the lower Pecos region a frequency greater than 2 percent in a coprolite suggests a strong possibility of intentional ingestion of flowers and certain seed types that still have pollen attached, and that a frequency of 10 percent should be interpreted as positive evidence of intentional ingestion.

Parasites and Nutrition

The presence of parasites observed in coprolites can help determine the amount of disease present in populations and indicate much about the subsistence and general quality of life. Examples include studies conducted by Henry J. Hall (1972) and Karl J. Reinhard (1985) in which differences were noted between the prevalence of parasitic disease in hunter-gatherers and in agriculturalists.

Agriculturalists and hunter-gatherers have very different subsistence bases and lifeways, which affect the types of diseases and parasites infecting each group. Hunter-gatherers were (and still are) mobile people who generally lived and moved in small groups and probably had limited contact with outsiders. They lived in temporary dwellings, usually moving in a seasonal pattern, and tended to enjoy a well-balanced diet (Dunn 1968). Subsistence centered on the environment and what it provided, making it the most important aspect of their existence (Nelson 1967; Hayden 1971).

Agriculturalists, by contrast, are generally sedentary and live in larger groups because of the increase in population that an agricultural subsistence base can support. Their dwellings are more permanent structures, and they have extensive contacts with groups because of a more complex society and because of extensive trading networks that link those societies.

Although population increase and sedentary agriculture seem linked, the tendency of sedentary agriculturalists to concentrate their diets largely on a single crop can adversely affect health, even though their numbers increase. Corn, for example, is known to be a poor source of iron and deficient in the essential amino acids lysine and tryptophan. Moreover, the phytic acid present in corn inhibits intestinal absorp-

tion of nutrients, all of which can lead to undernourishment and anemia (El-Najjar 1976; Walker 1985). Thus, the adoption of agriculture seems to have been accompanied by a decrease in nutritional status, although such a general proposition demands analysis in local or regional settings (Palkovich 1984; Rose et al. 1984).

Nutritional status, however, can also be affected by parasite infection, which sedentism tends to encourage (Nelson 1967). In the past, as people became more sedentary and population increased, human wastes were increasingly difficult to dispose of, poor sanitation methods increased the chances of food contamination, and water supplies were fouled (Walker 1985). Many parasites thrive in fecal material and create a breeding ground for disease. The problem is exacerbated as feces are used for fertilizer to produce larger crop yields. Irrigation is sometimes used to increase production, which promotes the proliferation of waterborne parasites and also aids the dispersal of bacteria (Cockburn 1967; Dunn 1968; Alland 1969; Fenner 1970; McNeill 1979).

In addition, as animals were domesticated they brought their own suite of parasites to the increasing pool of pathogens (Cockburn 1967; Alland 1969; Fenner 1970; McNeill 1979). And finally, the storage of grains and the disturbance of the local environment, which accompanies agricultural subsistence, can stimulate an increase in rodents and wild animals, respectively, and consequently an increase in their facultative parasites as well (Reinhard 1985).

It seems clear that the quality of the diet declined and parasite load increased as the transition was made to sedentary agriculture. This is not to say that hunter-gatherers were parasite-free. Rather, there was a parallel evolution of some parasites along with humankind's evolution from ancestral nonhuman primates to *Homo sapiens* (Kliks 1983). Then, as human mobility increased, some parasites were lost because of their specific habitat range and because of changes in environment and temperature. But as these latter changes occurred, new parasites were picked up. Probably such changes took place as humans migrated across the cold arctic environment of the Bering Strait into North America. Some parasites would have made the journey with their human hosts, whereas others were left behind (Cockburn 1967; McNeill 1979).

New Approaches to Dietary Reconstruction

Regional Syntheses

As more researchers are analyzing coprolites, regional syntheses of diet are becoming possible. Paul E. Minnis (1989), for example, has condensed a large coprolite data set from Anasazi populations in the Four Corners Region of the southwestern United States, covering a time period from Basketmaker III to Pueblo III (A.D. 500–1300). He observed that for the

sample area, local resource structure seemed to be more important for determining diet than chronological differences. For example, domesticated plants, particularly corn, were a consistent dietary item from Basketmaker III to Pueblo III; and there was a "generally stable dietary regime" during the time periods studied, although small-scale changes were also noted (Minnis 1989: 559).

In another example from the Lower Pecos Region of southwestern Texas and northern New Mexico, a total of 359 coprolite samples have been studied (Sobolik 1991a, 1994b). Analysis indicates that the prehistoric populations of the region relied on a wide variety of dietary items for their subsistence. A substantial amount of fiber was provided by the diet, particularly that derived from prickly pear, onion, and the desert succulents (agave, yucca, sotol). A large number of seed and nut types were also ingested, although prickly pear and grass seeds were the most frequent. Animal remains, especially bone and fur, were also observed with a high frequency in the coprolites, indicating that these prehistoric people were eating a variety of animals (e.g., rodents, fish, reptiles, birds, and rabbits). The ingestion of an extremely wide variety of flowers and inflorescences is also indicated by the coprolite pollen data.

Significant differences were also observed in the dietary components of the coprolites. These differences might be attributable to changes in dietary practice, particularly to an increase in the variety of the prehistoric diet. A more plausible explanation, however, is that such differences are a result of the different locations of the archaeological sites from which the coprolite samples were excavated.

These sites are located on a south-north and a west-east gradient. Sites located in the southwestern portion of the region are in a dryer, more desert environment (the Chihuahuan Desert) with little access to water, whereas the sites located in the northeastern portion of the region are closer to the more mesic Edwards Plateau, which contains a diversity of plants and trees and is close to a continuous water supply. Thus, dietary change reflected in the coprolites most likely represents spatial differences rather than temporal fluctuations (Sobolik 1991a).

Nutritional Analyses

Coprolites are extremely useful in providing dietary and nutritional data, although information from botanical, faunal, and human skeletal remains is also needed in any attempt to characterize the nutrition of a prehistoric population (Sobolik 1990, 1994a). A recent study involving the nutritional analysis of 49 coprolites from Nubia was conducted by Linda S. Cummings (1989). This analysis was unique in that the coprolites were taken from skeletal remains buried in cemeteries representing two distinct time periods, including the early Christian period (A.D. 550–750) and the late Christian period (up to A.D. 1450). It is

rare that prehistoric coprolites can actually be attributed to specific people, and the health of individuals can be assessed through both the coprolite remains and the human skeletal material, with one method illuminating the other.

In this case, the human skeletal material suggested that cribra orbitalia indicating anemia was the major indicator of nutritional stress. Coprolite analysis by Cummings (1989) revealed that there was probably a synergistic relationship in the diet of the population between iron-deficiency anemia and deficiencies of other nutrients, mainly folacin, vitamin C, vitamin B₆, and vitamin B₁₂. Cummings also noted differences in the diet and health of the two populations, including differences between males and females and older and younger members.

Pollen Concentration Studies

Although the determination of pollen concentration values has not been attempted in many coprolite studies, such values are important in determining which pollen types were most likely ingested. Studies show that after ingestion, pollen can be excreted for many days as the grains become caught in the intestinal folds. Experiments have also demonstrated that the concentration of intentionally ingested pollen can vary considerably in sequentially produced fecal samples (Kelso 1976; Williams-Dean 1978).

Glenna Williams-Dean (1978) conducted a modern fecal study that analyzed Brassicaceae and *Prosopis* pollen as a small component of pollen ingestion. It was revealed that Brassicaceae pollen was retained in the digestive system for much longer periods of time (up to one month after ingestion) than *Prosopis* pollen. Brassicaceae pollen is an extremely small grain with an average size of 12 micrometers (μm) and has a finely defined outer-wall sculpturing pattern. Both traits would most likely increase the retention of this pollen type in the folds of the intestine, allowing for it to be observed in many fecal samples. *Prosopis* pollen is a spherical, medium-sized grain (average size 30 μm), with a smooth exine sculpturing pattern. The larger size of this grain and the decreased resistance provided by the smooth exine would permit this pollen type to pass more quickly through the intestinal folds without retention.

In light of this study, it can be predicted that larger pollen grains, such as corn (*Zea*) and cactus (Cactaceae), and pollen with little exine sculpturing, such as juniper (*Juniperus*), will move quickly through the human digestive system. Thus, these pollen types would be observed in fewer sequential fecal samples than those of other types. By contrast, smaller pollen grains with significant exine sculpturing, such as sunflower pollen (high-spine Asteraceae), can be predicted to move more slowly through the digestive system, become frequently caught in the intestinal lumen, and thus observed in fecal samples many days after initial ingestion.

Such predictions were subsequently applied in an examination of prehistoric coprolites (Sobolik 1988). This investigation revealed that a high pollen concentration value in coprolite samples should indicate that the economic pollen types observed in the samples were ingested recently. Concentration values of over 100,000 pollen grains/gram of material usually contain recently ingested pollen. But samples that contain less than 100,000 pollen grains/gram of material may contain economic pollen types that were intentionally ingested many days before the sample was deposited (Sobolik 1988). Such samples will also contain a wide variety of unintentionally ingested, background pollen types. Therefore, it is more difficult to recognize intentionally ingested pollen types from samples that contain less than 100,000 pollen grains/gram of material.

Modern fecal studies are, thus, invaluable as guides in the interpretation of prehistoric coprolite pollen content and in indicating the limitations of the data. Many more such investigations are needed to determine both pollen percentage and concentration. The diet of the participants will have to be stringently regulated in order to minimize the influence of outside pollen contaminants, particularly those in bread and canned foods (Williams-Dean 1978). Ideally, such studies will include many people and, thus, as diverse a population of digestive systems as possible over a long period of time. An important addition to such a study would be the observation of the effect of a high-fiber and a high-meat diet on pollen output and fecal output in general.

Medicinal Plant Usage

Documenting prehistoric medicinal plant usage is problematic because it is difficult to distinguish between plants that were consumed for dietary purposes and those consumed for medicinal purposes. Indeed, in many instances plants were probably used both dietarily and medicinally.

Nonetheless, the analysis of plant remains from archaeological sites is often employed to suggest dietary and medicinal intake. Such remains can be deposited through a number of channels, most significantly by contamination from outside sources (i.e., water, wind, matrix shifts, and animals). Plants also were used prehistorically as clothing, shelter, baskets, and twining, and these, when deposited into archaeological contexts, can become mistaken for food or medicinal items. Here, then, is a reason why coprolites, which are a direct indication of diet, can provide new insights into prehistoric medicinal usage (Reinhard, Hamilton, and Hevly 1991).

In an analysis of the pollen content of 32 coprolites recovered from Caldwell Cave, Culberson County, Texas, Kristin D. Sobolik and Deborah J. Gerick (1992) revealed a direct correlation between the presence of plants useful for alleviating diarrhea and coprolites that were severely diarrhetic. This correlation suggests that the prehistoric population of Caldwell Cave was ingesting medicinal plants to help alleviate chronic

diarrhea. These plants, identified through analysis of the pollen content of the coprolites, included *Ephedra* (Mormon tea) and *Prosopis* (mesquite). Interestingly, this investigation confirmed the study conducted by Richard G. Holloway (1983), which indicated that the Caldwell Cave occupants were possibly using *Ephedra* and *Larrea* (creosote bush) in medicinal teas to help cure chronic diarrhea (Holloway 1983).

Mormon tea pollen, leaves, and stems are widely used as a diarrhetic and have been one of the most prevalent medicinal remedies for diarrhea both prehistorically and historically (Burlage 1968; Niethammer 1974; Moore 1979; Moerman 1986). Mesquite leaves are also useful as a medicinal tea for stomach ailments and to cleanse the digestive system (Niethammer 1974). As part of the process of preparing mesquite leaves for a medicinal tea, pollen could also become incorporated into the sample, either intentionally or unintentionally.

In another study, Reinhard and colleagues (1991) also determined medicinal plant usage through analysis of the pollen content in prehistoric coprolites. They found that willow (*Salix*), Mormon tea (*Ephedra*), and creosote (*Larrea*) were probably used for medicinal purposes prehistorically and that a large variety of other plants may have been used as well.

Protein Residues

It was previously mentioned that analysis of protein residues in coprolites is a relatively new advance in coprolite studies (Newman et al. 1993). This method attempts to link protein residues with the type of plant or animal that was consumed and involves the immunological analysis of tiny amounts of protein through crossover electrophoresis. The unknown protein residue from the coprolites is placed in agarose gel with known antiserum from different plants and animals. The agarose gel is then placed in an electrophoresis tank with a barbital buffer at pH 8.6, and the electrophoretic action causes the protein antigens to move toward the antibody that is not affected by the electrical action. The solution, which contains the unknown protein residue, and the matching plant or animal species antiserum form a precipitate that is easily identifiable when stained with Coomassie Blue R250 solution (Kooyman, Newman, and Ceri 1992). The samples that form a precipitate indicate that the matching plant or animal was eaten by the person who deposited the coprolite sample.

Two sample sets were selected for an analysis of the protein residues found in coprolites – seven from Lovelock Cave, Nevada, and five from an open site in the Coachella Valley of southern California (Newman et al. 1993). Protein analysis of samples from the open site was not successful. But the samples from Lovelock Cave indicated that human protein residues were found in six of the samples and protein residues from pronghorns were found in four samples. Such an initial study suggests that protein residue analysis can be

a successful and important component in the determination of prehistoric diet.

Gender Specificity

A new technique that will distinguish the gender of the coprolite depositor is presently being tested on coprolite samples from Mammoth and Salts Caves, Kentucky, by Patricia Whitten of Emory University. In this technique, which has been successful with primate studies, the gonadal (sex) steroids are removed from each coprolite sample and analyzed according to the content of testosterone, the male hormone, and estradiol, the female hormone. Both steroids can be found in each sample, but their frequencies vary depending upon gender. Modern human samples will first be analyzed to determine the frequencies of each gonadal steroid expected in males and females.

DNA analysis is also being attempted from coprolite samples to determine gender (Mark Q. Sutton 1994, personal communication). This type of research should allow archaeologists to determine dietary differences between males and females in a population and should also help in reconstructing the patterns of differential access to resources.

Conclusions

Reconstructing prehistoric human diets is a vast process requiring a variety of assemblages and disciplines to obtain a complete picture. Coprolite analysis provides a diverse and significant insight into the prehistoric diet. When analyzing the entire diet of a population, researchers must take into consideration information gleaned from other archaeological materials. Past coprolite research has focused on developing new and innovative techniques so that recovery of the diverse data inherent in such samples can be achieved. Such development has allowed researchers not only to observe the macrobotanical and macrofaunal remains from coprolites but also to analyze their pollen, parasite, and phytolith content.

Recent advances have seen the discipline progress toward determining medicinal plant ingestion so as to permit inter- and intraregional dietary comparisons; to determine the protein content of the samples; to analyze the nutritional content of the dietary items; and to determine the gender of the depositor of the sample. Coprolite analysis has definitely advanced out of its infancy, and its contributions to the determination of prehistoric diet, health, and nutrition in the future should prove to be significant indeed.

Kristin D. Sobolik

Bibliography

- Agassiz, L. 1833-43. *Recherches sur les poissons fossiles*. 5 vols. Vol. 2. Soleure. Neuchatel, Switzerland.
- Alland, A., Jr. 1969. Ecology and adaptation to parasitic diseases. In *Environmental and cultural behavior*, ed. A. P. Vayada, 115-41, Tucson.
- Bryant, Vaughn M., Jr. 1969. Late full-glacial and post-glacial pollen analysis of Texas sediments. Ph.D. dissertation, University of Texas.
- 1974a. Prehistoric diet in southwest Texas: The coprolite evidence. *American Antiquity* 39: 407-20.
- 1974b. The role of coprolite analysis in archeology. *Bulletin of the Texas Archeological Society* 45: 1-28.
1975. Pollen as an indicator of prehistoric diets in Coahuila, Mexico. *Bulletin of the Texas Archeological Society* 46: 87-106.
1986. Prehistoric diet: A case for coprolite analysis. In *Ancient Texans: Rock art and lifeways along the lower Pecos*, ed. Harry J. Shafer, 132-5. Austin.
1987. Pollen grains: The tiniest clues in archaeology. *Environment Southwest* 519: 10-13.
1990. Pollen: Nature's fingerprints of plants. *Yearbook of science and the future, Encyclopedia Britannica*. Chicago.
- Bryant, Vaughn M., Jr., and Glenna Williams-Dean. 1975. The coprolites of man. *Scientific American* 232: 100-9.
- Buckland, W. 1829. On the discovery of coprolites, or fossil faeces, in the Lias at Lyme Regis, and in other formations. *Geological Society of London, Transactions* 3: 223-36.
- Burlage, Henry M. 1968. *Index of plants of Texas with reputed medicinal and poisonous properties*. Austin, Tex.
- Callen, Eric O., and T. W. M. Cameron. 1960. A prehistoric diet revealed in coprolites. *The New Scientist* 8: 35-40.
- Carbone, Victor A., and B. C. Keel. 1985. Preservation of plant and animal remains. In *The analysis of prehistoric diets*, ed. R. I. Gilbert and J. H. Mielke, 1-19. New York.
- Cockburn, Aidan. 1967. *Infectious diseases: Their evolution and eradication*. Springfield, Ill.
- Cordell, Linda S. 1977. Late Anasazi farming and hunting strategies: One example of a problem in congruence. *American Antiquity* 42: 449-61.
- Cummings, Linda S. 1989. Coprolites from medieval Christian Nubia: An interpretation of diet and nutritional stress. Ph.D. dissertation, University of Colorado.
- Danielson, Dennis R. 1993. The role of phytoliths in prehistoric diet reconstruction and dental attrition. M.A. thesis, University of Nebraska.
- Davis, Owen K., Larry Agenbroad, Paul S. Martin, and J. I. Mead. 1984. The Pleistocene dung blanket of Bechan Cave, Utah. In *Contributions in quaternary vertebrate paleontology: A volume in memorial to John E. Guilday*, ed. H. H. Genoways and M. R. Dawson, 267-82. Pittsburgh, Pa.
- DeKay, J. E. 1830. On the discovery of coprolites in North America. *Philosophy Magazine* 7: 321-2.
- Douglas, Charles L. 1969. Analysis of hairs in Lovelock Cave coprolites. In *Archaeological and paleobiological investigations in Lovelock Cave, Nevada*, ed. L. K. Napton. Kroeber Anthropological Society Special Publication, 2: 1-8.
- Dunn, Frederick L. 1968. Epidemiological factors: Health and disease in hunter-gatherers. In *Man the hunter*, ed. R. B. Lee and I. DeVore, 221-8. Chicago.
- El-Najjar, M. 1976. Maize, malaria, and anemias in the New World. *Yearbook of Physical Anthropology* 20: 329-37.
- Faegri, K., and J. Iversen. 1964. *Textbook of pollen analysis*. New York.
- Fenner, Frank. 1970. The effects of changing social organization on the infectious diseases of man. In *The impact of civilization on the biology of man*, ed. S. V. Boyden, 71-82. Canberra, Australia.

- Ford, Richard I. 1979. Paleoethnobotany in American archaeology. In *Advances in archaeological method and theory*, Vol. 2, ed. Michael B. Schiffer, 285-336. New York.
- Fry, Gary F. 1985. Analysis of fecal material. In *The analysis of prehistoric diets*, ed. R. I. Gilbert and J. H. Mielke, 127-54. New York.
- Hall, Henry J. 1972. Diet and disease at Clyde's Cavern, Utah: As revealed via paleoscatology. M.A. thesis, University of Utah.
- Harshberger, J. W. 1896. The purpose of ethnobotany. *American Antiquarian* 17: 47-51.
- Hayden, Brian. 1971. Subsistence and ecological adaptations of modern hunter/gatherers. In *Omnivorous primates: Gathering and hunting in human evolution*, ed. R. S. O. Harding and G. Teleki, 154-65. New York.
- Hevly, Richard H., R. E. Kelley, G. A. Anderson, and S. J. Olsen. 1979. Comparative effects of climate change, cultural impact and volcanism in the paleoecology of Flagstaff, Arizona, A.D. 900-1300. In *Volcanic activity and human history*, ed. Payson D. Sheets and Donald K. Grayson, 120-8. New York.
- Hill, J., and Richard H. Hevly. 1968. Pollen at Broken K Pueblo: Some new interpretations. *American Antiquity* 33: 200-10.
- Holloway, Richard G. 1983. Diet and medicinal plant usage of a late Archaic population from Culberson County, Texas. *Bulletin of the Texas Archeological Society* 54: 27-49.
- Jennings, Jesse D. 1957. *Danger Cave*. University of Utah Anthropological Papers No. 27. Salt Lake City.
- Jones, Volney H. 1936. The vegetal remains of Newt Kash Hollow Shelter. *University of Kentucky Reports in Archaeology and Ethnology* 3: 147-65.
- Kelso, Gerald. 1976. Absolute pollen frequencies applied to the interpretation of human activities in Northern Arizona. Ph.D. dissertation, University of Arizona.
- Kirchoff, Paul. 1971. The hunting-gathering people of North Mexico. In *The north Mexican frontier*, ed. Basil C. Hedrick, J. Charles Kelley, and Carroll L. Riley, 200-9. Carbondale, Ill.
- Kliks, Michael. 1983. Paleoparasitology: On the origins and impact of human-helminth relationships. In *Human ecology and infectious diseases*, ed. N. A. Croll and J. H. Cross, 291-313. New York.
- Kooyman, B., M. E. Newman, and H. Ceri. 1992. Verifying the reliability of blood residue analysis on archaeological tools. *Journal of Archaeological Sciences* 19: 265-9.
- Laudermilk, J. D., and P. A. Munz. 1934. Plants in the dung of *Notbrotherium* from Gypsum Cave, Nevada. *Carnegie Institution of Washington Publication* 453: 29-37.
1938. Plants in the dung of *Notbrotherium* from Rampart and Muav Cave, Arizona. *Carnegie Institution of Washington Publication* 487: 271-81.
- Loud, L. L., and M. R. Harrington. 1929. *Lovelock Cave*. University of California Publications in American Archeology and Ethnology.
- Lyman, R. Lee. 1982. Archaeofaunas and Subsistence Studies. In *Advances in archaeological method and theory*, Vol. 5, ed. Michael B. Schiffer, 331-93. Tucson, Ariz.
- Mack, R. N., and Vaughn M. Bryant, Jr. 1974. Modern pollen spectra from the Columbia Basin, Washington. *North-west Science* 48: 183-94.
- MacNeish, Richard S. 1958. *Preliminary archeological investigations in the Sierra de Tamaulipas, Mexico*. American Philosophical Society Transactions, Vol. 44.
- Mantell, G. A. 1822. *The fossils of the South Downs; or illustrations of the geology of Sussex*. London.
- Martin, Paul S., B. E. Sables, and D. Shutler, Jr. 1961. Rampart Cave coprolite and ecology of the Shasta Ground Sloth. *American Journal of Science* 259: 102-27.
- Martin, Paul S., and F. W. Sharrock. 1964. Pollen analysis of prehistoric human feces: A new approach to ethnobotany. *American Antiquity* 30: 168-80.
- McNeill, William H. 1979. *The human condition: An ecological and historical view*. Princeton, N.J.
- Minnis, Paul E. 1989. Prehistoric diet in the northern Southwest: Macroplant remains from Four Corners feces. *American Antiquity* 54: 543-63.
- Moerman, Daniel E. 1986. *Medicinal plants of native America*. University of Michigan, Museum of Anthropology Technical Reports No. 19. Ann Arbor.
- Moore, Michael. 1979. *Medicinal plants of the mountain west*. Santa Fe, N. Mex.
- Munson, Patrick J., Paul W. Parmalee, and Richard A. Yarnell. 1971. Subsistence ecology at Scoville, a terminal Middle Woodland village. *American Antiquity* 36: 410-31.
- Napton, Lewis K. 1970. Archaeological investigation in Lovelock Cave, Nevada. Ph.D. dissertation, University of California, Berkeley.
- Napton, Lewis K., and O. A. Brunetti. 1969. Paleo-ornithology of Lovelock Cave coprolites. In *Archaeological and paleobiological investigations in Lovelock Cave, Nevada*, ed. Lewis K. Napton. Kroeber Anthropological Society Special Publication, 2: 9-18.
- Nelson, G. S. 1967. Human behavior in the transmission of parasitic diseases. In *Behavioural aspects of parasite transmission*, ed. E. U. Canning and C. A. Wright, 1-15. London.
- Newman, M. E., R. M. Yohe II, H. Ceri, and M. Q. Sutton. 1993. Immunological protein residue analysis of non-lithic archaeological materials. *Journal of Archaeological Science* 20: 93-100.
- Niethammer, Carolyn. 1974. *American Indian food and lore*. New York.
- Palkovich, A. M. 1984. Agriculture, marginal environments, and nutritional stress in the prehistoric southwest. In *Paleopathology at the origins of agriculture*, ed. M. N. Cohen and G. J. Armelagos, 425-38. New York.
- Parmalee, Paul W., Andreas Paloumpis, and Nancy Wilson. 1972. *Animals utilized by Woodland peoples occupying the Apple Creek Site, Illinois*. Illinois State Museum Reports of Investigations. Springfield.
- Reinhard, Karl J. 1985. Recovery of helminths from prehistoric feces: The cultural ecology of ancient parasitism. M.S. thesis, Northern Arizona University.
- Reinhard, Karl J., R. H. Brooks, S. Brooks, and Floyd B. Largent, Jr. 1989. Diet and environment determined from analysis of prehistoric coprolites from an archaeological site near Zape Chico, Durango, Mexico. *Journal of Paleopathology Monograph* 1: 151-7.
- Reinhard, Karl J., and Vaughn M. Bryant, Jr. 1992. Coprolite analysis: A biological perspective on archaeology. In *Advances in archaeological method and theory*, Vol. 4, ed. Michael B. Schiffer, 245-88. Tucson, Ariz.
- Reinhard, Karl J., Donny L. Hamilton, and Richard H. Hevly. 1991. Use of pollen concentration in paleopharmacology: Coprolite evidence of medicinal plants. *Journal of Ethnobiology* 11: 117-32.
- Robert, E. 1832-33. Sur des coprolithes trouvés à Passy. *Société de Géologie France Bulletin* 3: 72-3.
- Rose, Jerry C., B. A. Burnett, M. S. Nassaney, and M. W. Blaeuer. 1984. Paleopathology and the origins of maize agriculture in the lower Mississippi Valley and Caddoan culture areas. In *Paleopathology at the origins of agriculture*, ed. M. N. Cohen and G. J. Armelagos, 393-424. New York.
- Scott, Linda. 1987. Pollen analysis of hyena coprolites and sediments from Equus Cave, Taung, southern Kalahari

- (South Africa). *Quaternary Research, New York* 28: 144–56.
- Smith, Bruce D. 1975. *Middle Mississippi exploitation of animal populations*. University of Michigan, Museum of Anthropology Anthropological Papers No. 57. Ann Arbor.
- Smith, G. E., and F. W. Jones. 1910. *Archeological survey of Nubia 1907–1908*. Cairo.
- Sobolik, Kristin D. 1988. The importance of pollen concentration values from coprolites: An analysis of southwest Texas samples. *Palynology* 12: 201–14.
1990. A nutritional analysis of diet as revealed in prehistoric human coprolites. *Texas Journal of Science* 42: 23–36.
- 1991a. Paleonutrition of the Lower Pecos region of the Chihuahuan Desert. Ph.D. dissertation, Texas A&M University.
- 1991b. *The prehistoric diet and subsistence of the Lower Pecos region, as reflected in coprolites from Baker Cave, Val Verde County, Texas*. Studies in Archaeology Series No. 7, Texas Archaeological Research Lab, University of Texas, Austin.
1993. Direct evidence for the importance of small animals to prehistoric diets: A review of coprolite studies. *North American Archaeologist* 14: 227–44.
- 1994a. Introduction. In *Paleonutrition: The diet and health of prehistoric Americans*, ed. K. D. Sobolik, 1–18. Carbondale, Ill.
- 1994b. Paleonutrition of the Lower Pecos region of the Chihuahuan Desert. In *Paleonutrition: The diet and health of prehistoric Americans*, ed. K. D. Sobolik. Carbondale, Ill.
- Sobolik, Kristin D., and Deborah J. Gerick. 1992. Prehistoric medicinal plant usage: A case study from coprolites. *Journal of Ethnobiology* 12: 203–11.
- Spaulding, W. G. 1974. Pollen analysis of fossil dung of *Ovis canadensis* from southern Nevada. M.S. thesis, University of Arizona.
- Stiger, Mark A. 1977. Anasazi diet: The coprolite evidence. M.A. thesis, University of Colorado.
- Thompson, R. S., Thomas R. Van Devender, Paul S. Martin, et al. 1980. Shasta ground sloth *Nototheriops shastense* (Hoffsteter) at Shelter Cave, New Mexico; environment, diet, and extinction. *Quaternary Research, New York* 14: 360–76.
- Wakefield, E. F., and S. C. Dellinger. 1936. Diet of the Bluff Dwellers of the Ozark Mountains and its skeletal effects. *Annals of Internal Medicine* 9: 1412–18.
- Wales, S., J. Evans, and A. R. Leeds. 1991. The survival of waxes in coprolites: The archaeological potential. In *Archaeological sciences 1989: Proceedings of a conference on the application of scientific techniques to archaeology, Bradford, September 1989*, ed. P. Budd, B. Chapman, C. Jackson, et al. *Oxbow Monograph* 9, 340–4. Bradford, England.
- Walker, Phillip. 1985. Anemia among prehistoric Indians of the southwest. In *Health and disease of the prehistoric southwest*. Arizona State University Anthropological Research Papers No. 34, ed. C. F. Merbs and R. J. Miller, 139–64. Tempe.
- Webb, W. S., and R. S. Baby. 1957. *The Adena people*, no. 2. Columbus, Ohio.
- Williams-Dean, Glenna J. 1978. Ethnobotany and cultural ecology of prehistoric man in southwest Texas. Ph.D. dissertation, Texas A&M University.
- Young, B. H. 1910. *The prehistoric men of Kentucky*. Filson Club Publications No. 25. Louisville, Ky.

I.4. 🐾 Animals Used for Food in the Past: As Seen by Their Remains Excavated from Archaeological Sites

Animal remains excavated from archaeological sites are, to a large extent, the remnants of animals that were used for food. These remains include the fragmentary bones and teeth of vertebrates, the shells of mollusks, the tests of echinoderms, and the exoskeletal chitin of crustacea and insects. As with all archaeological remains, they represent discarded fragments of a previous way of life.

Organic remains are particularly subject to losses from the archaeological record, through the destructive nature of food preparation and consumption, the scavenging of refuse by other animals, and the deterioration that results from mechanical and chemical forces over time. Other losses come through excavation with inappropriate sieving strategies in which the remains of smaller individuals or species are lost. Nonetheless, despite all of these opportunities for the loss and destruction of organic material, animal remains constitute a major class of the archaeological remains from most sites and in some sites, such as shell mounds, they are the most obvious of the remains. However, even among those remains that are preserved, care must be taken in evaluating the extent to which they may represent a contribution to the prehistoric diet.

One reason for such caution is that all remains recovered are not necessarily those of animals that were consumed. For example, along the Gulf coast of Mexico, dogs were definitely eaten and probably even raised for food (Wing 1978). Their remains were often burned, disarticulated, and associated with those of other food remains. But in the West Indies, complete or nearly complete skeletons of dogs are found in burials and they are rarely associated with midden refuse, suggesting that dogs were not a regular item in the diet (Wing 1991). Dogs probably have played more roles in human culture than any other animal, ranging from food animal to guardian to hunting companion to faithful friend. But other animals, too, such as chickens, cattle, and horses, have likewise played different roles and thus their archaeological remains cannot automatically be assumed to constitute only the remnants of past meals.

Another problem is that some animals that were consumed left few remains. On the one hand, these were small, soft-bodied animals such as insect larvae, and on the other hand, they were very large animals, such as sea mammals or large land mammals that were too heavy to be brought back to the habitation site. In the case of small animals, mandibles of shrimp have recently been identified in some southeastern sites in the United States through the use of fine-gauge sieves (Quitmyer 1985). This find (which was

predicted) gives encouragement that other hard parts of otherwise soft-bodied animals can be found if they are carefully searched for. At the other end of the size scale, many very large animals were butchered at the kill site and only the meat was brought back to the home site, leaving little or no skeletal evidence at the latter site of this hunting enterprise.

Such a phenomenon has been termed the "schlepp effect" (Perkins and Daly 1968), which expresses the commonsense but nonetheless laborious practice of stripping the flesh off the carcass of a large animal and carrying it to the home site but leaving most of the heavy supporting tissue, the skeleton, at the kill site. Kill sites of single large prey species such as mammoths (*Mammuthus*) (Agenbroad 1984) and mass kills of bison (*Bison bison*) (Wheat 1972) have been excavated and the strategy of the kill and processing of the carcasses reconstructed. Good evidence points to the selection of prey based on its fatness and carcass use to maximize caloric intakes of the hunters (Speth 1983).

Human Adaptability

Human technology applied to food procurement and preparation is one of the factors responsible for the broad diet that sets humans apart from other animals and has made their worldwide distribution possible. Prehistoric sites with long sequences of occupation are located on every continent except Antarctica, and almost every island (Woodhouse and Woodhouse 1975). Such a wide distribution encompasses a great range of ecosystems with different potentials and limitations for human subsistence exploitation.

By contrasting the animal resources used in sites located in some of the major landforms, both the similarities and differences in the potentials of these ecosystems can be demonstrated. Some of the differences to be examined are between subsistence exploitation in sites located along rivers and on the marine coast. Other comparisons have to do with sites on continents as compared with those on islands, and subsistence in exploiting the arctic as opposed to the humid tropics.

Levels of Extraction

Different levels of extraction of animal foods from the environment obviously can affect the composition of the diet. The hunting, fishing, and gathering of resources are procurement enterprises that result in diverse food composition and variation throughout the year as different resources become available. During the great majority of human history, beginning several million years ago and extending to the time animals were controlled by domestication, the subsistence economy was based upon hunting, fishing, and gathering. Increased control over animals, whether through the maintenance of captive animals or managed hunting, culminated in animal domestication about 10,000 years ago. A comparison of economies

dependent on domestic animals with those relying on wild animals for subsistence reveals a range in diversity and dependability of resources.

The Nature of Animal Remains

Nature of Material

As already noted, the remains of animals used for food consist of the supporting tissue such as bone and teeth of vertebrates, shell of mollusks, and chitin of crustaceans. These tissues are composed of inorganic and organic compounds. The relatively rigid structure of the inorganic portions predominate in bone, constituting 65 percent. In tooth enamel the inorganic portion is 99.5 percent by weight. The inorganic portions of these supporting tissues are composed of compounds of calcium.

By their very nature, skeletal remains are attractive to scavengers: Some meat and other soft tissue may have adhered to them and, in addition, bone itself is sought by many animals as a source of calcium.

Other losses of archaeological remains come about through the influence of forces of nature, termed taphonomic factors. Such natural changes include all types of site disturbances such as erosion or stream washing, land movement, alternating freezing and thawing, and acidic soil conditions. The soil conditions are particularly critical for the preservation of bone, which as a calcium compound can be dissolved in acidic conditions. Destruction of bone under acidic conditions is further complicated by the problem of bone loss being greatest in the least well calcified bones of young individuals. In contrast, losses are the smallest in the enamel portion of teeth (Gordon and Buikstra 1981). Alternating freezing and thawing is the other taphonomic factor that is particularly damaging to organic remains. Bones exposed to the sun develop cracks. When these cracks fill with moisture and then freeze, they enlarge and, ultimately, the bone will fragment into pieces that have lost their diagnostic characteristics.

If losses from the archaeological record can complicate reconstruction of the past, so too can additions to the faunal assemblage. Such additions were often animals that lived and died at the habitation site. These are known as commensal animals, the most well known being the black and the Norway rats (*Rattus rattus* and *Rattus norvegicus*) and the house mouse (*Mus musculus*). Burrowing animals such as moles (*Talpidae*) and pocket gophers (*Geomyidae*) may also dig into a site and become entombed. In addition, middens and habitation sites were occasionally used for burial, and the remains of both people and their animals (i.e., dogs) were thus inserted into earlier deposits.

Other ways in which commensal animals can be incorporated in a midden is by association with the target species. For example, many small creatures, such as mussels, snails, crabs, and barnacles, adhere to

clumps of oysters and were often brought to the site on this target species. Occasionally, too, the stomach contents of a target species may contain remains of other animals, which are thus incorporated in the site.

Recovery and Identification

Optimum recovery of faunal material is clearly essential for reconstruction of past diets. In some cases, sieving the archaeological remains with 7-millimeter (mm)-gauge screens will be sufficient to recover the animal remains. But as more and more sieving experiments are conducted and archaeologists gain more and more experience using fine-gauge (2 mm and 4 mm) sieves, it is becoming increasingly obvious that fine-gauge sieves are essential for optimal recovery at most sites. A good example of this importance has to do with what was long thought to be the enigma of the preceramic monumental site of El Paraiso on the Pacific coast of Peru. This site was an enigma because early excavations there uncovered no vertebrate remains, suggesting that the local ancient residents had no animal protein in their diets. It was only when sieving with 2-millimeter-gauge sieves was undertaken that thousands of anchovy (*Engraulidae*) vertebrae were revealed. At last, the part this small schooling fish played in providing fuel for the construction of the monumental architecture was understood (Quilter et al. 1991).

Another methodological problem that needs further consideration at this site and at others is the relative contribution of vertebrates and invertebrates to the prehistoric diet. Mollusk shells are relatively more durable and less likely to be lost to scavengers than are the bones of vertebrates of similar size, which results in a bias favoring mollusks.

In evaluating the potential contribution of vertebrates and mollusks to the prehistoric diet, one must also keep in mind biases in their preservation and recovery. For example, molluskan shells are more massive relative to their edible soft tissues than are the skeletons of vertebrates to their soft tissues. Consequently, shells will be the most visible component of a shell midden even before taking into account the greater durability of shell and the fewer losses to scavengers.

One way some of these problems can be addressed is by making estimations of potential meat weight using allometric relationships between measurable dimensions of the shell or bone and meat weight. Such relationships, of course, can only be developed by using the accurate weights and measurements of modern specimens comparable to the species recovered from archaeological deposits. Knowledge of modern animals and a reference collection are the most important research tools of the faunal analyst.

First, an attempt is made to estimate the minimum weight of the meat represented by the animal remains. Next, this estimate is contrasted with the maximum estimate of the meat weight that could

have been provided by the minimum number of individual animals calculated to be represented. If the two estimates are approximately the same, this implies that all of the meat from each animal was consumed at the source of the refuse. Yet if the two estimates differ substantially, it suggests that only portions of some animals were consumed at the site, with other portions distributed within the community.

The Meaning of Animal Remains Assemblages

Animal Exploitation in Different Ecosystems

Riverine versus coastal. Many similarities exist between faunal assemblages from riverine sites and those located along the coast. In each case, both vertebrate and invertebrate aquatic animals were important food sources. Furthermore, these two aquatic situations were subject to influxes of animals, either in breeding congregations or migrations, that augmented the resident fauna. Of course, aquatic species used by riverine fishermen and gatherers were different from those extracted from the sea.

Aquatic organisms, fishes and mollusks, are important in the faunal assemblages of both riverine and coastal sites. Shell mounds are more typically associated with coastal sites, although they do occur along rivers (Weselkov 1987). Riverine shell mounds are accumulations typical of the Archaic time period (about 7000 to 3000 B.P.) in eastern North America, exemplified by those found on the St. Johns River in Florida and the Green River in Kentucky (Claassen 1986). Along coastal shores, shell mounds date from Archaic times to the present.

Gathering mollusks is invariably accompanied by fishing, but the relative contribution of fish and shellfish to the diet follows no pattern. Many shell mounds are visually very impressive and, in fact, are so large that in many places these mounds have been mined for shell material to be used in modern road construction. Less visible components of these archaeological sites are the vertebrate, crustacean, and material cultural remains.

As indicated in the discussion of methods, those for reconstructing the dietary contributions of the two major components of a shell mound are still being perfected. One exciting development has been a greater understanding of the importance (as food) of very small animals in the vertebrate, predominantly fish, components of these mounds.

As already mentioned, the use of fine-gauge screen sieves in the recovery of faunal remains has provided a more accurate understanding of the size range of fishes consumed in the past. Catches of small fishes are being documented in many parts of the world. For example, at four sites in the Darling Region of Australia, otoliths of golden perch (*Maguaria ambigua*) were preserved. These could be used to extrapolate lengths of fishes caught up to 24,000 years ago (Balme 1983). The range in their estimated lengths is

8 to 50 centimeters (cm) and the mean is approximately 20 cm. The range in estimated lengths of sardines (*Sardina pilchardus*) from a fourth-century A.D. Roman amphora is between 11 and 18 cm (Wheeler and Locker 1985). Catfish (*Ariopsis felis*) and pinfish (*Lagodon rhomboides*) from an Archaic site on the southeastern Gulf coast of Florida are estimated to range in length from 4 cm to 25 cm, and the means of the catfish and pinfish are 10 cm and 12 cm, respectively (Russo 1991).

An important question has to do with how these small fishes could be eaten and yet leave skeletal remains behind. Many contemporary diets include small fishes such as sardines and anchovies in which the entire body of the fish is consumed. The answer may lie in the fact that small fishes are used in many parts of the world in the preparation of fish sauces that are made without skeletal tissue. In other words, the well-preserved, intact skeletal remains of small fishes suggest that the fishes might have been employed in sauces.

In addition to mollusks and fishes in the protein portion of a coastal and riverine diet, migratory or breeding congregations would have added significantly to the diet. A well-known example of this phenomenon is the seasonal migration of anadromous fishes such as salmon (*Salmonidae*) and herring (*Clupeidae*) (Rostlund 1952); methods of preserving and storing this seasonal surplus would also have developed.

Other examples of such exploitation include the breeding colonies of sea bird rookeries, sea turtle nesting beaches, and seal and sea lion colonies. Many of these colonies are coastal phenomena and are strictly confined to particular localities. During the breeding cycle, most animals are particularly vulnerable to predation, and people through the ages have taken advantage of this state to capture breeding adults, newborn young, and, in the cases of birds and turtles, eggs.

Unfortunately, only some of the evidence for this exploitation can be demonstrated in the archaeological remains. Egg shells are rarely preserved. Some of the breeding animals in question, like the sea mammals and the sea turtles, are very large and may have been butchered on the beach and the meat distributed throughout the community. Thus, it is difficult to assess the relative importance of these resources within a particular refuse deposit and by extension within the diet of the humans being studied.

Continental versus island. A continental fauna differs from an island fauna in its diversity of species. There is a direct relationship between island size and distance from the mainland and species diversity (MacArthur and Wilson 1967). Human exploitation on a continent can range from catches of very diverse species in regions where different habitats are closely packed, to dependence on one or two species that form herds,

typically in open grasslands. On islands other than very large ones, prehistoric colonists found few species and often augmented what they did find with the introduction of domestic species as well as tame captive animals.

This kind of expansion seems to be a pattern in many parts of the world. For example, several marsupials (*Phalanger orientalis*, *Spilococus maculatus*, and *Thylogale brunii*), in addition to pigs (*Sus scrofa*) and dogs (*Canis familiaris*), were deliberately introduced into the Melanesian Islands between 10,000 and 20,000 years ago (Flannery and White 1991). Similarly, sheep (*Ovis* sp.), goats (*Capra* sp.), pigs (*Sus* sp.), and cats (*Felis* sp.) were all introduced into the Mediterranean Islands at a time when the domestication of animals was still in its initial stages. A variety of wild animals, such as hares (*Lepus europaeus*), dormice (*Glis glis*), foxes (*Vulpes vulpes*), and badgers (*Meles meles*), were also introduced into this area (Groves 1989).

Likewise, in the Caribbean Islands, domestic dogs as well as captive agouti (*Dasyprocta leporina*), opossum (*Didelphis marsupialis*), and armadillo (*Dasyus novemcinctus*) were introduced from the South American mainland, whereas the endemic hystriognath rodent locally called "hutia" (*Isolobodon portoricensis*) and an endemic insectivore (*Nesophontes edithae*) were introduced from large to small islands (Wing 1989).

Although tame animals were doubtless kept by people living on the mainland, they are not easily distinguished from their wild counterparts. But this problem is only part of an increasingly complex picture as human modifications of the environment, either overtly through landscape changes resulting from land clearing or more subtly through hunting pressure, have altered the available species on both islands and continental land masses.

Because of the generally lower species diversity on islands, exploitation of terrestrial species was augmented by marine resources. These were primarily fishes and mollusks. In the Caribbean Islands, West Indian top shell (*Cittarium pica*) and conch (*Strombus gigas*), and a whole array of reef fishes including parrotfishes (Scaridae), surgeonfishes (*Acanthuridae*), grouper (Serranidae), and jacks (Carangidae), were of particular importance.

Arctic versus humid tropics. The Arctic has long, cold, dark winters but also short summers, with long spans of daylight that stimulate a brief period of extraordinarily high plant productivity. By contrast, the humid tropics have substantially more even temperatures and lengths of daylight throughout a year that in many cases is punctuated by dry and rainy seasons. Needless to say, these very different environmental parameters have a pronounced effect on the animal populations available for human subsistence within them.

Traditional contemporary subsistence activities as

well as evidence from archaeological faunal remains in the Alaskan Arctic indicate that important contributors to the human diet have been caribou (*Rangifer tarandus*), sea mammals – particularly seals (*Calorhinus ursinus* and *Phoca vitulina*) and sea lions (*Eumatopius jubata*) – seabirds, and marine fishes, primarily cod (*Gadus macrocephalus*) (Denniston 1972; Binford 1978; Yesner 1988).

Although the species of animals differ in different parts of the Arctic regions, many characteristics of a northern subsistence pertain. Foremost among these is a marked seasonal aspect to animal exploitation correlated with the migratory and breeding patterns of the Arctic fauna. Moreover, the length of time during which important animal resources, such as the anadromous fishes, are available becomes increasingly circumscribed with increased latitude (Schalk 1977). To take full advantage of this glut of perishable food, people need some means of storage. And fortunately, in a region where temperatures are below freezing for much of the year, nature provides much of the means.

Another characteristic of northern subsistence is the generally low species diversity; but this condition is counteracted by some large aggregations of individuals within the species. Still, the result is that heavy dependence is placed on a few species. At Ashishik Point an analysis of the food contribution of different animals to the prehistoric diet revealed that sea mammals and fishes predominated (Denniston 1972). Of the sea mammals, sea lions are estimated consistently to have provided the greatest number of calories and edible meat. This estimation agrees with the observation by David Yesner (1988) that dietary preference among the prehistoric Aleut hunter-gatherers was for the larger, fattier species.

The fauna of the humid tropics is much more diverse than that of the Arctic, although the tropical animals that people have used for food do not generally form the large aggregations seen in higher latitudes. Exceptions include schools of fishes, some mollusks, bird rookeries, and sea turtle breeding beaches. Many of the animals that have been exploited are relatively small. The largest animals from archaeological sites in the New World tropics are adult sea turtles (*Cheloniidae*), deer (*Mazama americana* and *Odocoileus virginianus*), and peccary (*Tayassu pecari* and *Tayassu tajacu*) (Linares and Ranere 1980). Some of the South American hystricognath rodents, such as the capybara (*Hydrochaeris* sp.), paca (*Agouti paca*), and agouti (*Dasyprocta punctata*), were used and continue to be used widely as food. Fish and shellfish were also very important components of prehistoric diets (Linares and Ranere 1980) and typically augmented those based on terrestrial plants and animals.

The kinds of land mammals that have been most frequently consumed in much of the tropics since the beginning of sedentary agriculture prompted Olga Linares (1976) to hypothesize a strategy for their capture she called “garden hunting.” She suggested that

many of these animals were attracted to the food available in cultivated fields and gardens and were killed by farmers protecting their crops. Objective ways of evaluating the importance of garden hunting have been proposed, which are based on the composition of the faunal assemblage (Neusius 1996).

Different Levels of Extraction

Hunting, fishing, and gathering. It has been estimated that fully 90 percent of all those who have lived on earth have done so as hunter-gatherers (Davis 1987). The animals that were procured by these individuals varied depending upon the resources available within easy access of the home site or the migratory route. But in addition to these differences in the species that were obtained and consumed, certain constraints governed what was caught, by whom, how the meat was distributed, and how it was prepared. Certainly, technology played an important part in what kinds of animals were caught on a regular basis and how the meat was prepared. Many specialized tools such as water craft, nets, traps, weirs, bows, arrows, and blowguns and darts all permitted the capture of diverse prey.

A limitation imposed on all organisms is that of energy, which meant that food procurement generally took place within what has become known as the “catchment” area (Higgs and Vita-Finzi 1972). This area is considered to be inside the boundaries that would mark a two-hour or so one-way trip to the food source. Theoretically, travel more distant than this would have required more energy than could be procured. Once animal power was harnessed, trips to a food source could be extended. However, another solution to the problem of procurement of distant resources was by adopting a mobile way of life. This might have taken the form of periodic trips from the home site or it might have been a migratory as opposed to a sedentary method of living.

In the past, as today, there was doubtless a division of labor in the food quest. It is generally thought that men did the hunting and fishing (although in many traditional societies today women do the inland, freshwater fishing) whereas the women, children, and older male members of the community did the gathering. Some excellent studies of contemporary hunter-gatherers provide models for these notions about the division of labor. One of the best and most frequently cited is by Betty Meehan (1982) entitled *Shell Bed to Shell Midden*, in which she describes shellfishing practices in detail and the relative importance of shellfish to the diet throughout the year. This is the case at least among the Gidjingali-speaking people of Arnham Land in northern Australia, whose shellfish gathering is a planned enterprise entailing a division of labor for collecting molluscan species.

Food sharing is another phenomenon that probably has great antiquity in the food quest, although admittedly, we know more about the patterns of food shar-

ing from contemporary studies than from archaeological remains. A classic study of sea turtle fishing along the Caribbean coast of Nicaragua (Nietschmann 1973) describes the distribution obligations of meat obtained from the large sea turtle (*Chelonia midas*). Such patterns make certain that meat does not go to waste in the absence of refrigeration and furthermore assure the maintenance of community members who are less able or unable to procure food for themselves.

That a large carcass was shared can sometimes be detected in an archaeological assemblage of animal remains. As we noted earlier, this observation can occur when estimates of meat yield based on the actual remains recovered are compared with estimates of potential meat obtained from the calculated numbers of individual animals represented in the same sample. Disparity between these two estimates could point to the incomplete deposit of the remains of a carcass, which may indicate sharing, either among the households, or through a distribution network, or even through a market system. Plots of the dispersal of parts of deer carcasses throughout a prehistoric community that was entirely excavated provide a demonstration of food distribution (Zeder and Arter 1996). Perhaps sharing also occurred when many small fishes were caught in a cooperative effort. In this case the catch may have been shared equally with an additional share going to the owner of the fishing equipment.

Communal cooperation was probably also involved in the concept of "garden hunting" (Linares 1976), which can be viewed in the broader perspective of deliberately attracting animals. Food growing in gardens or fields, or stored in granaries, was used as bait to entice animals to come close to a settlement.

Clearly this strategy would have been most successful with animals that ate agricultural products. Some of those so trapped were probably tamed and eventually domesticated, suggesting that it is no accident that many of our domestic and tame animals consume crop plants. Today agriculture and animal husbandry are combined to produce the plant and animal foods consumed throughout most of the world. But the cultivation of crops and the husbandry of animals did not arise simultaneously everywhere. In much of the Western Hemisphere crops were grown in the absence of a domestic animal other than the dog. However, the management, control, and domestication of animals eventually led to a different level of exploitation.

Animal domestication. Domestic animals have skeletal elements and teeth that are morphologically distinct from their wild ancestors. The observable changes are the result of human selection, which is why domestic animals are sometimes referred to as man-made animals. Human selection prior to modern animal husbandry may have been unintentional and more a result of isolation and methods of confinement (e.g., whether animals were tethered, or kept in stalls or corrals).

Animals were, of course, tamed first, and paintings on walls and on pottery sometimes provide evidence of domestication. Selection expressed as changes in the morphology of an animal also indicates domestication, but the state of animal "tamelessness" is difficult to recognize in the fragmentary remains from archaeological sites and consequently rarely possible to document. Moreover, many animals were held captive and tamed but for some reason never became domesticated, meaning that their skeletal remains do not differ morphologically from their wild counterpart. Collared peccaries (*Tayassu tajacu*), for example, were believed to have been tamed and kept by the Maya for food or ritual purposes (Hamblin 1984), and stone pens found on the island of Cozumel are thought to have been used to keep peccaries for human convenience. At the present time, peccaries are trained by some hunters in Central America to fill the role of watchdogs. Clearly, many motives instigated the taming of animals, but the most important of these undoubtedly was ready access to meat.

Some animals were held in captivity with no effort made to tame them. The captivity of these animals was a means of storing fresh meat. An example of animals still kept this way today are sea turtles maintained in corrals along the shore below low tide. Live animals were also maintained on board ships during long ocean voyages. These practices probably were very old, and animals such as domestic pigs were often released on islands to assure a source of food for the next voyage or for other people who followed.

But control over animals, or even their domestication, did not necessarily mean a sole reliance on them for food. Rather, especially in the early stages of domestication, humans continued to depend on hunted and fished resources. But with the introduction of livestock into new regions, traditional subsistence strategies were modified. An interesting example may be seen in sixteenth-century Spanish Florida, where Spanish colonists, suddenly confronted with wilderness and relative isolation, changed their traditional subsistence in a number of ways. They abandoned many accustomed food resources and substituted wild resources used by the aboriginal people of the region. Moreover, their animal husbandry practices shifted from a traditional set of domesticated animals to just those that flourished in the new environment (Reitz and Scarry 1985). These changes required flexibility. Yet presumably, such changes in subsistence behavior documented for Spanish settlers in Florida were repeated in many places throughout the ages (Reitz and Scarry 1985).

Dependence upon livestock became more complete only after peoples and their agricultural systems were well established. It should be noted, however, that even in industrial societies today, reliance upon domestic animals is not complete. Most contemporary Western diets include fish and other seafood, and

wild animal food, such as venison, is viewed as a delicacy and is often the food for a feast.

Accompanying a greater human dependence upon domestic animals for food was an increased human use of the animals' energy and other products. Animals, used for draft, greatly increased the efficiency of agricultural enterprises, and utilizing them for transportation extended the catchment area.

The employment of animals for dairy products was of crucial importance and can be detected in archaeological remains by the kill-off pattern that characteristically shows male individuals killed as young animals and females maintained to old age for their milk (Payne 1973). When such provisions as milk (and eggs) came into use, they provided an edible resource without loss of the animal, and the latter came to be viewed as capital.

Conclusion

The human diet is characterized by the great variety of plants and animals that it includes. The animal protein portion of the diet depends most heavily on vertebrates and mollusks but also includes crustaceans, arthropods, and echinoderms. Historically, this flexibility has been instrumental in the worldwide distribution of human beings.

Despite this flexibility, selection for certain resources was clearly practiced. Usually, certain species were targeted even though a great variety of species were used. When selection was exercised, the determining factors seemed to have been those of a dependable resource and one high in body fat. By resource dependability we mean, for example, the annual salmon run, stable oyster beds, perhaps a captive flock of pigeons or a domestic herd of goats. Selection for animals with the highest body fat has been documented specifically in the preference for sea lions by the prehistoric Aleuts and generally by archaeological remains of the bison, which revealed selection of the fattest individuals. In this connection it should be noted that domestic animals tend to store more fatty tissue than do their wild ancestors, which may have been a further incentive for the maintenance of domesticates (Armitage 1986).

Food distribution and sharing is another characteristic of human subsistence that provided everyone with a degree of security even if personal catches were not successful. Methods of food preparation and storage doubtless varied greatly. Most likely these were salting, smoking, or drying, or a combination, but none of these methods is clearly visible in archaeological remains. It is only when animal remains are found far outside of their normal range (e.g., codfish remains in West Indian historic sites) that one can be sure some method of preservation was employed.

Similarly, cooking methods can be interpreted from archaeological remains only in a general way. It is likely that meat was boiled or stewed when bone was not burned and roasted when it was burned.

Yet such interpretations must be made cautiously because bone can become burned in many ways.

Even though animal remains from archaeological sites are fragmentary and much is often missing from the whole picture, they do provide an important perspective on past human diets.

Elizabeth S. Wing

Bibliography

- Agenbroad, L. D. 1984. New World mammoth distribution. In *Quaternary extinctions: A prehistoric revolution*, ed. P. S. Martin and R. G. Klein, 90–108. Tucson, Ariz.
- Armitage, Philip L. 1986. Domestication of animals. In *Bio-industrial ecosystems*, ed. D. J. A. Cole and G. C. Brender, 5–30. Amsterdam.
- Balme, Jane. 1983. Prehistoric fishing in the lower Darling, western New South Wales. In *Animals and archaeology*, ed. C. Grigson and J. Clutton-Brock, 19–32. Oxford.
- Binford, Lewis R. 1978. *Nunamuit ethnoarchaeology*. New York.
- Claasen, Cheryl. 1986. Shellfishing seasons in the prehistoric southeastern United States. *American Antiquity* 51: 21–37.
- Davis, Simon J. M. 1987. *The archaeology of animals*. New Haven, Conn.
- Denniston, Glenda B. 1972. Ashishik point: An economic analysis of a prehistoric Aleutian community. Ph.D. dissertation, University of Wisconsin.
- Flannery, T. F., and J. P. White. 1991. Animal translocation. *National Geographic Research and Exploration* 7: 96–113.
- Gordon, Claire C., and Jane E. Buikstra. 1981. Soil pH, bone preservation, and sampling bias at mortuary sites. *American Antiquity* 46: 566–71.
- Groves, Colin P. 1989. Feral mammals of the Mediterranean islands: Documents of early domestication. In *The walking larder*, ed. J. Clutton-Brock, 46–58. London.
- Hamblin, Nancy L. 1984. *Animal use by the Cozumel Maya*. Tucson, Ariz.
- Higgs, E. S., and C. Vita-Finzi. 1972. Prehistoric economies: A territorial approach. In *Papers on economic prehistory*, ed. E. S. Higgs, 27–36. London.
- Linares, Olga F. 1976. "Garden hunting" in the American tropics. *Human Ecology* 4: 331–49.
- Linares, Olga F., and Anthony J. Ranere, eds. 1980. *Adaptive radiations in prehistoric Panama*. Peabody Museum Monographs No. 5. Cambridge, Mass.
- MacArthur, Robert H., and Edward O. Wilson. 1967. *The theory of island biogeography*. Princeton, N.J.
- Meehan, Betty. 1982. *Shell bed to shell midden*. Canberra, Australia.
- Neusius, S. W. 1996. Game procurement among temperate horticulturalists: The case for garden hunting by the Dolores Anasazi. In *Case studies in environmental archaeology*, ed. E. J. Reitz, L. A. Newsom, and S. J. Scudder, 273–88. New York.
- Nietschmann, Bernard. 1973. *Between land and water: The subsistence ecology of the Miskito Indians*. New York.
- Payne, Sebastian. 1973. Kill-off patterns in sheep and goats: The mandibles from Asvan Kale. *Anatolian Studies* 23: 281–303.
- Perkins, Dexter, Jr., and Patricia Daly. 1968. A hunter's village in Neolithic Turkey. *Scientific American* 219: 96–106.
- Quilter, Jeffery, Bernardino E. Ojeda, Deborah M. Pearsall, et

- al. 1991. Subsistence economy of El Paraiso, an early Peruvian site. *Science* 251: 277-83.
- Quitmyer, Irvy R. 1985. The environment of the Kings Bay locality. In *Aboriginal subsistence and settlement archaeology of the Kings Bay locality*, Vol. 2, ed. W. H. Adams, 1-32. University of Florida, Reports of Investigations No. 2. Gainesville.
- Reitz, Elizabeth J., and C. Margaret Scarry. 1985. *Reconstructing historic subsistence with an example from sixteenth-century Spanish Florida*. Society for Historical Archaeology, Special Publication No. 3. Glassboro, N.J.
- Rostlund, Erhard. 1952. *Freshwater fish and fishing in North America*. University of California Publications in Geography, Vol. 9. Berkeley.
- Russo, Michael. 1991. Archaic sedentism on the Florida coast: A case study from Horr's Island. Ph.D. dissertation, University of Florida.
- Schalk, Randall F. 1977. The structure of an anadromous fish resource. In *For theory building in archaeology*, ed. L. R. Binford, 207-49. New York.
- Speth, John D. 1983. *Bison kills and bone count*. Chicago.
- Weselkov, Gregory A. 1987. Shellfish gathering and shell midden archaeology. In *Advances in archaeological method and theory*, Vol. 10, ed. M. B. Schiffer, 93-210. San Diego, Calif.
- Wheat, Joe Ben. 1972. The Olsen-Chubbuck site: A Paleo-Indian bison kill. *American Antiquity* 27: 1-180.
- Wheeler, Alwyne, and Alison Locker. 1985. The estimated size of sardines (*Sardina pilchardus*) from amphorae in a wreck at Randello, Sicily. *Journal of Archaeological Science* 12: 97-100.
- Wing, Elizabeth S. 1978. Use of dogs for food: An adaptation to the coastal environment. In *Prehistoric coastal adaptations*, ed. B. L. Stark and B. Voorhies, 29-41. New York.
1989. Human exploitation of animal resources in the Caribbean. In *Biogeography of the West Indies*, ed. C. A. Woods, 137-52. Gainesville, Fla.
1991. Dog remains from the Sorcé Site on Vieques Island, Puerto Rico. In *Beavers, bobwhites, and blue-points*, ed. J. R. Purdue, W. E. Klippel, and B. W. Styles, 389-96. Springfield, Ill.
- Woodhouse, David, and Ruth Woodhouse. 1975. *Archaeological atlas of the world*. New York.
- Yesner, David R. 1988. Effects of prehistoric human exploitation on Aleutian sea mammal populations. *Arctic Anthropology* 25: 28-43.
- Zeder, M. A., and S. R. Arter. 1996. Meat consumption and bone use in a Mississippian village. In *Case studies in environmental archaeology*, ed. E. J. Reitz, L. A. Newsum, and S. J. Scudder, 319-37. New York.

1.5. Chemical Approaches to Dietary Representation

Dietary reconstruction for past populations holds significant interest as it relates to biological and cultural adaptation, stability, and change. Although archaeological recovery of floral and faunal remains within a prehistoric or historical context provides some direct evidence of the presence (and sometimes quantity) of potential food resources, indirect evidence for the

dietary significance of such foodstuffs frequently must be deduced from other bioarchaeological data.

The types of data with dietary significance range from recovered plant and animal remains through evidence of pathology associated with diet, growth disruption patterns, and coprolite contents. Other traditional approaches involving the people themselves – as represented by skeletal remains – include demographic (Buikstra and Mielke 1985) and metabolic (Gilbert 1985) stress patterns.

In addition to bioanthropological analyses, reconstruction of environmental factors and the availability and limits of food species and their distribution for a population with a particular size, technology, and subsistence base are typical components within an archaeological reconstruction. Although these physical aspects are significant, the distribution, or more likely the restriction, of particular foodstuffs from certain segments of the population (because of sex, age, status, food avoidance, or food taboos) may be important cultural system features. The seasonal availability of food and its procurement, preservation, and preparation may also have influenced group dietary patterns and nutritional status (Wing and Brown 1979).

Analysis of skeletal remains may also provide some direct evidence of diet. Type and adequacy of diet have long been of interest to physical anthropologists, especially osteologists and paleopathologists (Gilbert and Mielke 1985; Larsen 1987). More recently, direct chemical analysis of bones and teeth has been attempted in an effort to assess the body's metabolism and storage of nutritive minerals and other elements. L. L. Klepinger (1984) has reviewed the potential application of this approach for nutritional assessment and summarized the early findings reported in the anthropological literature. (In addition, see Volume 14 of the *Journal of Human Evolution* [1985], which contains significant research surveys to that date.)

General Approaches and Assumptions

The pioneering anthropological work in bone chemical analysis and dietary reconstruction can be attributed to A. B. Brown (1973), who examined strontium concentrations in relation to meat and vegetation, and to R. I. Gilbert (1975), who explored the concentrations of five other elements in relation to prehistoric Native American samples in Illinois. The enthusiastic early promise of a relatively easy, straightforward approach to diet reconstruction from elemental concentrations in bone has more recently been tempered by recognition of the biodynamic complexity, methodological problems, and contextual changes occurring through diagenesis of buried bone. Nonetheless, a number of publications in article, dissertation, and book form have appeared in the anthropological literature from the early 1970s to the current time. Although the emphasis, samples, and time frame have varied considerably, the approaches generally share

the assumptions that bone is the vital feature for mineral homeostasis and a reservoir for critical elements, and that variations in bone concentrations by individual and group reflect past intakes of dietary concentrations, which in turn reflect local environmental and cultural milieus. A. C. Aufderheide (1989) and M. K. Sandford (1992, 1993a) have provided excellent reviews of the basic premises, biogenic-diagenetic continuum, diversity of methods, and sampling and analytical protocols. Two recent, extremely important, edited volumes (Price 1989; Sandford 1993b) contain the most comprehensive bibliographies currently available and include syntheses of recent findings, remaining problems, and potential research trajectories and protocols.

Initial anthropological bone chemical research focused primarily on the inorganic mineral phase of bone – which typically makes up 75 to 80 percent of the dry weight – and was concerned with the dietary contrasts and trophic levels of early hominids, hunter-gatherers, and agriculturalists. Analysis of the stable isotopes in the organic collagen component (20 to 25 percent) is now frequently undertaken to investigate the relative importance of C3/C4 plants and reliance on maize in the Americas through a consideration of its carbon content (Bumstead 1984). Such analysis is also focused on the equally troublesome problem of the marine/terrestrial protein components of the diet by investigation of nitrogen isotopes. Other isotopes with possible relevance to aspects of dietary reconstruction include those of oxygen, sulfur, and strontium. W. F. Keegan (1989), M. A. Katzenberg (1992), and S. H. Ambrose (1993) provide excellent introductory reviews of the use of isotopes in the analysis of prehistoric diet. H. P. Schwarcz and M. J. Schoeninger (1991) provide a more esoteric review of the theory and technical details of isotopic analysis for reconstructing human nutritional ecology.

Samples, Instrumentation, and Variables

It is probably unrealistic to expect a finely detailed reconstruction of past diets from skeletal chemical data because of the nature of our evidence. A number of factors influence the survival and recovery of human remains. These include mortuary method, climatic conditions, soil chemistry, decomposition rates, and archaeological methods and goals. Although a single individual, may reflect important aspects of the biocultural past, including diet, we should remember that each individual, and the circumstances and context of the recovery of that individual, is unique.

Moreover, because the bone is analyzed not in life, but after death, bone turnover rates must be viewed in relation to age and health status at the time of that death. For aggregate data, especially for statistical comparisons, the representativeness, comparable size, and composition of the samples are also of concern. Chemical analysis should be done only after thorough

and professional osteological analysis is conducted. Useful standardized guides in such analysis are the Paleopathology Association's *Skeletal Database Committee Recommendations* (1991) and the collective recommendations for standard skeletal data collection (Buikstra and Ubelaker 1994).

Accurate demographic profiles are especially important for later considerations of age, sex, health/disease, and perhaps status categories within the population sample. Individual bone samples may be taken for later analysis, especially now that many invaluable skeletal collections are being reburied. Because as little as 1 to 2 grams of bone may be used for chemical analysis, depending upon the instrumentation and method, the removal and laboratory destruction of this amount of bone represents a loss of even less than that of a typical tooth.

Chemical concentrations within bone vary from one bone to another and even in different portions of an individual bone; understandably, earlier comparative studies were difficult before this fact was recognized. Current recommendations are to use cortical bone, preferably from femur midshafts. The particular technique chosen for quantitative elemental analysis will depend upon the number of elements to be analyzed, the cost, and the degree of precision required. Aufderheide (1989) and Sandford (1992) review the theoretical foundations and relative advantage of a number of options. Current laboratory analytical techniques include electroanalysis, light spectrometry, scanning electron microscopy, neutron activation analysis, mass spectrometry, and the widely used inductively coupled plasma (ICP) optical spectrometry for multiple element analysis. Results are typically reported in parts per million of bone or bone ash, the latter preferred (Price et al. 1989).

Although elemental analysis may be conducted directly on bone or in solution, isotopic analysis first requires decalcification, extraction of collagen from bone – approximately 5 milligrams (mg) of collagen is needed – and then analysis through mass spectrometry. Katzenberg (1992) provides a capsular review of the process and cites B. S. Chisholm (1989) and Schoeninger and colleagues (1989) as current basic references. Stringent laboratory conditions and lengthy preparation techniques, frequently with elevated costs, are necessary for isotopic analysis (Ambrose 1990).

Diagenesis appears to be less of a problem with isotopic analysis; however, it must still be considered. Much recent research in elemental analysis has attempted to document and cope with the problems of chemical and physical changes that may occur in buried bone through leaching, contextual contamination, and chemical reactions of bone outside the living body. In addition to the bone, samples of the matrix of soil must be collected and analyzed so that potential contamination may be identified.

Of course, postmortem influences must be detected, but physiological processes may also influence the

incorporation and utilization of elements present in a particular diet. Absorption of particular elements ingested may be enhanced or reduced through chemical processes or physiological regulation by competing substances in the diet. Phytates found in some cereals, for example, may bind zinc and iron and reduce the absorption of these elements.

In addition, not all elements are distributed equally through body tissues, and some, such as lead or strontium, may be deposited differentially into bone. Metabolism differences for particular elements may also compound the interpretation. Retention of ingested elements is variable as well in certain tissues as is the rate of bone turnover at different ages and under variable health conditions.

Finally, excretion rates for particular elements depend upon physiological processes and the nature of the element itself. Although there are a great number of variables in the incorporation, retention, and analysis of any particular chemical element in bone, a cautious application of trace element studies with appropriate samples, methods, and situations or research goals continues to hold promise for some aspects of dietary reconstruction. Indeed, despite severe critical evaluation of earlier trace element studies (Radosevich 1993), improved and refined multidisciplinary research and laboratory protocols should prevent the necessity of throwing the baby out with the bathwater.

Anthropological Dietary Chemical Reconstructions

The following sampling of past contributions within anthropological bone chemical research reflects three major thrusts related to diet: trophic level, temporal change in subsistence, and distinctive chemical elements. The general trophic level of past human diets was first investigated by strontium and strontium/calcium ratios. The basic premise was that the relative reliance on meat derived from animals higher on the food chain would be reflected by a lower concentration of strontium in the human bone because of its differential presence and absorption (relative to calcium) along the food chain. For paleoanthropologists concerned with the physical and cultural development of humans and the hunting complex (especially *Australopithecus*, *Homo habilis*, and *Homo erectus*), it first appeared that the answers could be derived in a relatively straightforward manner (Sillen and Kavanagh 1982). However, the fossilization process itself appears to alter the initial concentration, and other diagenic processes may confound the interpretation (Sillen, Sealy, and van der Merwe 1989).

In the case of more recent human groups, however, an analysis of strontium bone content has proven more fruitful. The anthropological significance of the strontium content of human bone was initially investigated by Brown (1973), and subsequent studies

of strontium content suggest that dietary differences may reflect social stratification. Schoeninger (1979), for example, determined that at a prehistoric Mexican site, higher-ranking individuals – as indicated by interment with more grave goods – had lower levels of strontium and, hence, presumably a greater access to animal protein. Other studies, such as that by A.A. Geidel (1982), appear to confirm the value of strontium analysis in this respect, even though diagenetic change must be evaluated.

Temporal dietary changes and the relative amounts of meat and plants in the diet (perhaps related to population size as well as technological complexity) have been documented in a number of regions. Gilbert (1975, 1977), for Late Woodland Mississippian groups in the Midwest, and T. D. Price and M. Kavanagh (1982), for the same area, document an increasing strontium concentration among groups as they experienced an increasing reliance on cereals and a concomitant decrease in meat availability. Katzenberg (1984) determined similar temporal changes among Canadian groups, as did Schoeninger (1981) for areas of the Middle East.

It should be noted, however, that bone strontium concentrations are strongly influenced by ingestion of marine foods – such as shellfish – and some nuts. J. H. Burton and Price (1990) suggest that low barium/strontium ratios distinguish consumption of marine resources. It should also be noted that soil and water concentrations of strontium and, hence, plant absorption of it, also vary geographically. A final caveat is the documentation of the influences of physiological processes such as weaning (Sillen and Smith 1984) and pregnancy and lactation (Blakely 1989), which elevate bone strontium and depress maternal bone calcium concentrations.

A number of other elements found in food and water (Ca, Na, Sr, Cu, Fe, Mn, Mg, Zn, Al, Fe, Ba) have the potential for assisting in dietary reconstruction. These elements have been analyzed in skeletal samples with varying degrees of success in delineating food categories, temporal changes, and subsample variations related to age, gender, or class. Like strontium, these elements are subject to many of the same modifications and processes from ingestion to deposition into bone, and frequently to the same diagenic processes after death, so the same caveats apply to their analysis and interpretation.

In addition, when these various elements are incorporated together in diets (and later deposited in bone), they may be antagonistic to each other, or enhanced when ingested together or as part of the same diet. In anthropological analysis, although the major elements such as calcium or phosphorous may be significant, the majority of research has been concerned with trace elements in either their total concentration for dietary categories, or as deficiencies related to particular diseases, or at toxic levels, such as lead poisoning.

Although there is a relatively abundant literature on individual trace elements and their role in human metabolism and nutrition in the medical and nutrition literature (Underwood 1977; Prasad 1978; Rennert and Chan 1984), these studies tend to focus on Western diets and modern food standards and samples. The major emphasis within anthropological elemental studies has been with meat and vegetable dietary questions and temporal change, especially in the prehistoric American Southeast and Middle West. Research in other world areas has included Europe (Grupe and Herrmann 1988), Southwest Asia (Sillen and Smith 1984), Sicily (Klepinger, Kuhn, and Williams 1986), Tunisia (Sandford, Repke, and Earle 1988), Australia (Kyle 1986), and Peru (Edward and Benfer 1993).

The theoretical premise behind such investigations is based on the different concentration levels of particular elements in dietary resources that then should vary in the human skeletal concentrations. Meat, for example, is typically associated with increased concentrations of iron, zinc, copper, molybdenum, and selenium. Plants, however, generally have greater amounts of strontium, magnesium, manganese, cobalt, and nickel. Unfortunately, a single plant or animal species rarely possesses a unique chemical signature. Besides the problem of mixed dietary resources, many of the prevailing trace elements overlap (Gilbert 1985), and nuts present special problems (Buikstra et al. 1989). Synthetic critical reviews of relevant literature have been provided by Price (1989), Sandford (1992, 1993a, 1993b), Aufderheide (1989), and J. E. Buikstra and colleagues (1989).

The emerging consensus is that elemental and isotopic studies may indeed be significant in circumstantial dietary reconstructions of past populations. But additional research is necessary to cope with the numerous problems and issues connected with such studies. Among these are diagenesis, laboratory analysis and sample preparation, expansion to skeletal samples of more recent origin, wider geographical representations and inclusions, feeding experiments, and more sophisticated statistical and interpretative techniques.

A number of studies have attempted to deal with the question of diagenesis and the need for adjustments before statistical analysis (Lambert, Szpunar, and Buikstra 1979; Katzenberg 1984; Price 1989; Edward and Benfer 1993; Radosevich 1993). Multiple bone analyses, comparisons with nonhuman animals (herbivores, carnivores, and mixed feeders), more multielement surveys, and careful laboratory evaluation are recommended.

Expansion of multielement or single element studies into more recent historical periods should have the advantage of combining available historic information concerning diet and food habits with the chemical analysis of skeletal samples for a more comprehensive understanding. For example, Aufderheide and colleagues (1981, 1985, 1988) have delineated socioeco-

nomic differences, occupational categories, and probably food storage patterns from the analysis of skeletal lead in the United States colonial period. H. A. Waldron (1981, 1983) and T. Waldron (1982, 1987) have addressed similar problems in the United Kingdom. In like fashion, J. S. Handler, Aufderheide, and R. S. Corrucini (1986) combined nineteenth-century descriptions of "dry bellyache" among Barbados slaves with an analysis of slave remains to demonstrate that "dry bellyache" was actually lead poisoning, the result of contaminated rum from stills with lead fittings. T. A. Rathbun and J. D. Scurry (1991) found regional variation in lead burdens in skeletal samples of whites and blacks from the seventeenth- and eighteenth-century eastern United States. Such variations seem to reflect differences in socioeconomic class, food preparation, and drinking patterns. Whites, who made far greater use of drinking and eating utensils, carried considerably higher lead burdens than blacks, with those of the Middle Atlantic states having slightly higher levels than those of other Southeast samples. Among blacks, females had the highest levels, indicating that they also doubtless had greater access to the whites' lead-contaminated food and drink.

Utilizing techniques of chemical analysis, W. D. Wood, K. R. Burns, and S. R. Lee (1985) and Rathbun (1987) were able to document regional and perhaps cultural differences among rural blacks, plantation slaves, and white elites in the nineteenth-century southeastern United States. Among the findings were that whites apparently had more access to meat than did either enslaved or freed African-Americans.

Similarly, Rathbun (1987) and T. A. J. Crist (1991) found dietary variation by gender, age, and perhaps stress level among a nineteenth-century South Carolina plantation slave sample. Males seem to have relied more heavily on meats, grains, and nuts than females, whose diets consisted more of leafy and leguminous vegetables. The remains of older adults reflected diets of grains, vegetables, and seafood, whereas those of younger adults revealed the consumption of more meats and, perhaps, nuts. Analysis of historical documents concerning food allocations on the plantation suggests that much of this differential was because slaves supplemented plantation rations with food items they collected and cooked themselves. Clearly, in many instances, a combining of historical, anthropological, and chemical information has the potential for providing a richer determination of past dietary contents and the consequences of various dietary regimens.

Summary

In addition to the confounding problems of preservation, diagenesis, data collection, and analysis, if elemental and isotopic chemical analysis of skeletal material is to fulfill its potential in dietary reconstruction, insightful and appropriate avenues of interpretation are necessary. Although descriptive statistics of aggre-

gate data drawn from a sample are useful heuristic devices, selection of appropriate analytical techniques appear to be linked to the nature of the concentration distributions.

Parametric and nonparametric – as well as univariate and multivariate – statistics have been applied to bone chemical quantitative data. The multiple problems and considerations involved have recently been discussed by Buikstra and colleagues (1989), who ultimately recommend principal component analysis. Even though mathematical rigor remains extremely important, insightful interpretations of relationships and findings still seem to require evaluation within a biocultural context.

Schoeninger (1989), for example, attempted to match food component elements and isotopes as well as skeletal analysis for prehistoric Pecos Pueblo and historic Dutch whalers to propose reasonable diets for them. Klepinger (1992) also commented on the importance of reevaluating model hypotheses that are frequently invoked in the light of new data and developing technologies.

Chemical approaches to dietary representation, especially of past groups, can be fascinating, frustrating, and fulfilling. But it seems unlikely that we will soon develop a comprehensive picture of past diets through chemical analysis alone. The complexity of the geochemical, biochemical, biological, physiological, cultural, and social systems involved require collaborative research and multidisciplinary sharing of results. Although each discipline and researcher may contribute various pieces of the puzzle, a clearer image can emerge only through integrative interpretations. The goal appears well worth the effort!

Ted A. Rathbun

Bibliography

- Ambrose, S. H. 1990. Preparation and characterization of bone and tooth collagen for isotopic analysis. *Journal of Archaeological Science* 17: 431–51.
1993. Isotopic analysis of paleodiets: Methodological and interpretive considerations. In *Investigations of ancient human tissue: Chemical analyses in anthropology*, ed. M. K. Sandford, 59–130. Langhorne, Pa.
- Aufderheide, A. C. 1989. Chemical analysis of skeletal remains. In *Reconstruction of life from the skeleton*, ed. M. Y. Iscan and K. A. R. Kennedy, 237–60. New York.
- Aufderheide, A. C., J. L. Angel, J. O. Kelley, et al. 1985. Lead in bone III: Prediction of social content in four Colonial American populations (Catoclin Furnace, College Landing, Governor's Land and Irene Mound). *American Journal of Physical Anthropology* 66: 353–61.
- Aufderheide, A. C., F. D. Neiman, L. E. Wittmers, and G. Rapp. 1981. Lead in bone II: Skeletal lead content as an indicator of lifetime lead ingestion and the social correlates in an archaeological population. *American Journal of Physical Anthropology* 55: 285–91.
- Aufderheide, A. C., L. E. Wittmers, G. Rapp, and J. Wallgren. 1988. Anthropological applications of skeletal lead analysis. *American Anthropologist* 90: 932–6.
- Blakely, R. E. 1989. Bone strontium in pregnant and lactating females from archaeological samples. *American Journal of Physical Anthropology* 80: 173–85.
- Brown, A. B. 1973. Bone strontium content as a dietary indicator in human skeletal populations. Ph.D. dissertation, University of Michigan.
- Buikstra, J. E., S. Frankenberg, J. Lambert, and L. Xue. 1989. Multiple elements: Multiple expectations. In *The chemistry of prehistoric human bone*, ed. T. D. Price, 155–210. Cambridge.
- Buikstra, J. E., and J. H. Mielke. 1985. Demography, diet, and health. In *The analysis of prehistoric diets*, ed. R. I. Gilbert, Jr., and J. H. Mielke, 360–422. Orlando, Fla.
- Buikstra, J. E., and D. H. Ubelaker. 1994. *Standards for data collection from human skeletal remains*. Arkansas Archaeological Survey, Research Series No. 44. Fayetteville.
- Bumstead, M. P. 1984. *Human variation: ^{13}C in adult bone collagen and the relation to diet in an isochronous C4 (Maize) archaeological population*. Los Alamos, N. Mex.
- Burton, J. H., and T. D. Price. 1990. Ratio of barium to strontium as a paleodietary indicator of consumption of marine resources. *Journal of Archaeological Science* 17: 547–57.
- Chisholm, B. S. 1989. Variation in diet reconstructions based on stable carbon isotopic evidence. In *The chemistry of prehistoric human bone*, ed. T. D. Price, 10–37. Cambridge.
- Crist, T. A. J. 1991. *The bone chemical analysis and bioarchaeology of an historic South Carolina African-American cemetery*. Volumes in Historical Archaeology XVIII. Columbia, S.C.
- Edward, J. B., and R. A. Benfer. 1993. The effects of diagenesis on the Paloma skeletal material. In *Investigations of ancient human tissue: Chemical analyses in anthropology*, ed. M. K. Sandford, 183–268. Langhorne, Pa.
- Geidel, A. A. 1982. Trace element studies from Mississippian skeletal remains: Findings from neutron activation analysis. *MASCA Journal* 2: 13–16.
- Gilbert, R. I., Jr. 1975. Trace element analysis of three skeletal Amerindian populations at Dickson Mounds. Ph.D. dissertation, University of Massachusetts.
1977. Applications of trace element research to problems in archaeology. In *Biocultural adaptations to prehistoric America*, ed. R. I. Blakely, 85–100. Athens, Ga.
1985. Stress, paleonutrition, and trace elements. In *The analysis of prehistoric diets*, ed. R. I. Gilbert, Jr., and J. H. Mielke, 339–58. Orlando Fla.
- Gilbert, R. I., Jr., and J. H. Mielke, eds. 1985. *The analysis of prehistoric diets*. Orlando, Fla.
- Grupe, G., and B. Herrmann. 1988. *Trace elements in environmental history*. Heidelberg.
- Handler, J. S., A. C. Aufderheide, and R. S. Corruccini. 1986. Lead content and poisoning in Barbados slaves. *Social Science History* 10: 399–425.
- Katzenberg, M. A. 1984. *Chemical analysis of prehistoric human bone from five temporally distinct populations in Southern Ontario*. Ottawa.
1992. Advances in stable isotope analysis of prehistoric bones. In *Skeletal biology of past peoples: Research methods*, ed. S. R. Saunders and M. A. Katzenberg, 105–19. New York.
- Keegan, W. F. 1989. Stable isotope analysis of prehistoric diet. In *Reconstruction of life from the skeleton*, ed. M. Y. Iscan and K. A. R. Kennedy, 223–36. New York.
- Klepinger, L. L. 1984. Nutritional assessment from bone. *Annual Review of Anthropology* 13: 75–9.
1992. Innovative approaches to the study of past human health and subsistence strategies. In *Skeletal biology of*

- past peoples: Research methods*, ed. S. R. Saunders and M. A. Katzenberg, 121–30. New York.
- Klepinger, L. L., J. K. Kuhn, and W. S. Williams. 1986. An elemental analysis of archaeological bone from Sicily as a test of predictability of diagenetic change. *American Journal of Physical Anthropology* 70: 325–31.
- Kyle, J. H. 1986. Effect of post-burial contamination on the concentrations of major and minor elements in human bones and teeth. *Journal of Archaeological Science* 13: 403–16.
- Lambert, J. B., C. B. Szpunar, and J. E. Buikstra. 1979. Chemical analysis of excavated human bone from middle and late Woodland sites. *Archaeometry* 21: 403–16.
- Larsen, C. S. 1987. Bioarchaeological interpretations of subsistence economy and behavior from human skeletal remains. *Advances in Archaeological Method and Theory* 10: 339–445.
- Paleopathology Association. 1991. *Skeletal database committee recommendations*. Detroit, Mich.
- Prasad, A. S. 1978. *Trace elements and iron in human metabolism*. New York.
- Price, T. D., ed. 1989. *The chemistry of prehistoric human bone*. Cambridge.
- Price, T. D., G. J. Armelagos, J. E. Buikstra, et al. 1989. The chemistry of prehistoric human bone: Recommendations and directions for future study. In *The chemistry of prehistoric human bone*, ed. T. D. Price, 245–52. Cambridge.
- Price, T. D., and M. Kavanagh. 1982. Bone composition and the reconstruction of diet: Examples from the midwestern United States. *Midcontinent Journal of Archaeology* 7: 61–79.
- Radosevich, S. C. 1993. The six deadly sins of trace element analysis: A case of wishful thinking in science. In *Investigations of ancient human tissue: Chemical analyses in anthropology*, ed. M. K. Sandford, 269–332. Langhorne, Pa.
- Rathbun, T. A. 1987. Health and disease at a South Carolina plantation: 1840–1870. *American Journal of Physical Anthropology* 74: 239–53.
- Rathbun, T. A., and J. D. Scurry. 1991. Status and health in colonial South Carolina: Belleview plantation, 1738–1756. In *What mean these bones?: Studies in southeastern bioarchaeology*, ed. J. L. Powell, P. S. Bridges, and A. M. W. Mires, 148–64. Tuscaloosa, Ala.
- Rennert, O. M., and W. Chan. 1984. *Metabolism of trace metals in man*. Boca Raton, Fla.
- Sandford, M. K. 1992. A reconsideration of trace element analysis in prehistoric bone. In *Skeletal biology of past peoples: Research methods*, ed. S. R. Saunders and M. A. Katzenberg, 79–103. New York.
- 1993a. Understanding the biogenic-diagenetic continuum: Interpreting elemental concentrations of archaeological bone. In *Investigations of ancient human tissue: Chemical analyses in anthropology*, ed. M. K. Sandford, 3–57. Philadelphia, Pa.
- ed. 1993b. *Investigations of ancient human tissue: Chemical analyses in anthropology*. Philadelphia, Pa.
- Sandford, M. K., D. B. Repke, and A. L. Earle. 1988. Elemental analysis of human bone from Carthage: A pilot study. In *The circus and a Byzantine cemetery at Carthage*, ed. J. H. Humphrey, 285–96. Ann Arbor, Mich.
- Schoeninger, M. J. 1979. Diet and status at Chalcatzingo: Some empirical and technical aspects of strontium analysis. *American Journal of Physical Anthropology* 51: 295–310.
1981. The agricultural “revolution”: Its effect on human diet in prehistoric Iran and Israel. *Paleorient* 7: 73–92.
1989. Reconstructing prehistoric human diet. In *The chemistry of prehistoric human bone*, ed. T. D. Price, 38–67. Cambridge.
- Schoeninger, M. J., K. M. Moore, M. K. Murray, and J. D. Kingston. 1989. Detection of bone preservation in archaeological and fossil samples. *Applied Geochemistry* 4: 281–92.
- Schwarcz, H. P., and M. J. Schoeninger. 1991. Stable isotope analyses in human nutritional ecology. *Yearbook of Physical Anthropology* 34: 283–321.
- Sillen, A., and M. Kavanagh. 1982. Strontium and paleodietary research: A review. *Yearbook of Physical Anthropology* 25: 67–90.
- Sillen, A., J. C. Sealy, and N. J. van der Merwe. 1989. Chemistry and paleodietary research: No more easy answers. *American Antiquity* 54: 504–12.
- Sillen, A., and P. Smith. 1984. Sr/Ca ratios in juvenile skeletons portray weaning practices in a medieval Arab population. *Journal of Archaeological Science* 11: 237–45.
- Underwood, E. J. 1977. *Trace elements in human and animal nutrition*. New York.
- Waldron, H. A. 1981. Postmortem absorption of lead by the skeleton. *American Journal of Physical Anthropology* 55: 395–8.
1983. On the postmortem accumulation of lead by skeletal tissues. *Journal of Archaeological Science* 10: 35–40.
- Waldron, T. 1982. Human bone lead concentrations. In *Roman-British cemeteries at Cirencester*, ed. A. McWhirr, L. Viner, and C. Wells, 203–7. Gloucester, England.
1987. The potential of analysis of chemical constituents of bone. In *Death, decay and reconstructions: Approaches to archaeology and forensic science*, ed. A. Boddington, A. N. Garland, and R. C. Janaway, 149–59. Manchester, England.
- Wing, E. S., and A. B. Brown. 1979. *Paleonutrition: Method and theory in prehistoric foodways*. New York.
- Wood, W. D., K. R. Burns, and S. R. Lee. 1985. *The Mt. Gilead cemetery study: An example of biocultural analysis from western Georgia*. Athens, Ga.

I.6. 🐾 History, Diet, and Hunter-Gatherers

In the years since 1960 there has been a dramatic change in our perception of the diet, nutrition, and health of “hunter-gatherers,” who constitute the world’s smallest, most “primitive,” and presumably oldest-style societies. The Hobbesian perspective (Hobbes 1950, original 1651), which assumes that malnutrition, disease, and hardship characterize primitive life – a view that prevailed among scholars for the nineteenth and the first half of the twentieth centuries – has been challenged during recent decades by a large series of new observations and a new theoretical paradigm.

Contemporary Hunter-Gatherers

Studies of African hunter-gatherers by Richard Lee (1968, see also 1969) and James Woodburn (1968), in the influential anthology *Man the Hunter* (Lee and DeVore 1968), suggested that far from living on the

edge of starvation, primitive hunter-gatherers frequently enjoyed not only adequate and well-balanced nutrition but also a relatively light workload.

In his analysis of the diet and workload of the !Kung San hunter-gatherers of the Kalahari Desert in southern Africa, Lee (1968, 1969) noted that the San diet consisted of an eclectic, yet selective, collection of wild foods – mostly (about 80 percent) vegetable, eaten fresh. He found that the San consumed 23 of 85 plant species that they knew to be edible in their environment and 17 of 55 edible animal species.

He calculated that for a relatively small investment of time, San hunter-gatherers obtained an adequate and well-balanced diet. By obtaining chemical analyses of their native foods and estimating the quantity of each food consumed by every individual, he was able to show that theoretically, each individual in the group received sufficient protein, vitamins, and minerals. In contrast to modern diets, what seemed the “limiting” factor – the element in the San diet most likely to be short or lacking – was the number of calories it delivered. Lee estimated the caloric intake at about 2,140 kilocalories (kcal) per person per day during a season of the year that he considered neither the richest nor the poorest.

Similarly, Woodburn (1968), although less precise, was even more sanguine in his description of the diets of the Hadza of Tanzania, who frequented a far richer environment than that of the !Kung San. He described their quest for food as leisurely and richly rewarding.

Medical observations of the San (Truswell and Hansen 1976) confirmed that they showed no signs of qualitative malnutrition, in that they had no visible deficiencies of vitamins, minerals, or protein, although they may have been showing signs of low caloric intake. (Low calorie intake has been cited by various others as responsible for stunting San growth and reducing their fertility.) Also of note was an absence of high blood pressure and elevated serum cholesterol, as well as the scarcity of heart problems. (See also Bronte-Stewart et al. 1960; Wehmeyer, Lee, and Whiting 1969; Metz, Hart, and Harpending 1971). Particularly striking were the observations on both the San and the Hadza suggesting that children did not suffer from the kinds of childhood malnutrition – kwashiorkor, marasmus, and associated weanling diarrhea – that were otherwise so common in African children (Jelliffe et al. 1962; Truswell and Hansen 1976).

At about the same time that these studies were emerging, agricultural economist Ester Boserup (1965) proposed a new model of economic growth in human history. She argued that population growth rather than technological progress had been the main stimulus for economic change. “Primitive” behavior, although usually considered to be a function of ignorance, might, she suggested, be seen as an efficient adjustment to a small population and a small social scale. So-called progress, she argued, might simply be a necessary adjustment to increasing population size,

scale, and density and might be associated with declining rather than improving labor efficiency and declining rather than improving individual welfare.

Based on the work of Boserup, Woodburn, and Lee, a number of archaeologists proposed that the initial adoption of farming by prehistoric hunting and gathering groups, which occurred in various parts of the world beginning about 10,000 years ago (the “Neolithic Revolution”), might also have been a grudging response to ecological stress or population “pressure” on resources. In other words, the Neolithic Revolution might not have been the result of technological progress as had previously been assumed (Binford 1968; Flannery 1969; Cohen 1977).

One of these scholars (Cohen 1977) extended the argument by suggesting that much of what had passed for progress in prehistory might, like the adoption of farming, have been a response to the pressure of growing population, rather than the result of new inventions, since the new “progressive” techniques seemed to represent the input of extra effort for relatively little output. These “improvements” would include the adoption of diets based on small seeds and the development of grindstones to process them; the development of small projectiles for hunting small game; and the increase in shellfish consumption and concomitant development of fishing equipment during the Mesolithic or Archaic stages of prehistoric economic development. It is true that an argument could be mounted that such apparent economic trends may be distorted by problems of archaeological preservation. For example, the scarcity of shellfish remains in earlier prehistory might reflect poor preservation. However, it is difficult to defend a similar argument about the late appearance of grindstones and small projectile points.

Questions and Challenges for the New Perspectives

A number of questions remain about these new perspectives and the data upon which they were originally developed. For example, it is not clear whether the !Kung San are as well nourished and as affluent as Lee presented them (see also Sahlins 1968). Nor is it clear that the !Kung San are typical of modern hunter-gatherers in the quality of their nutrition. And finally, it is not clear that they, or *any* contemporary hunter-gatherers, lead lives that are representative of the historic and prehistoric experience of human hunters.

In the matter of the nutritional state of the !Kung San, G. Silberbauer (1981) has suggested that the groups of San he studied were nutritionally depressed and might have been lacking in B vitamins. Similarly, Edwin Wilmsen (1978) estimated that San caloric intake might fall well below 2,000 kcal per person in the poorest season. Others such as Kristen Hawkes and J. F. O’Connell (1985) and N. Blurton-Jones and P. M. Sibley (1978) have also argued that the San are not as well nourished as they have been described,

that their caloric intake may be deficient, and that their "leisure" time may actually be an adjustment to the extreme heat and dryness of the Kalahari, which limits activity for significant portions of the year.

Moreover, Carmel Schrire (1980, 1984) and others have questioned the value of the !Kung San and other contemporary hunter-gatherers as models for prehistory, arguing that they are not remnants of an ancient way of life but, rather, modern societies formed by contemporary political and economic conditions in South Africa and elsewhere. As such, according to Schrire, their experience has little meaning for the study of prehistory.

Some New Evidence

Recent work in several fields has suggested that the broad perspectives introduced by Lee, Woodburn, and Boserup are accurate even though some details of their arguments may be open to challenge (Cohen 1989).

Such work rests, at least in part, on the assumption that despite undeniable pressures and inputs from the larger societies that surround them, contemporary hunter-gatherer societies can (with appropriate caution) be viewed as twentieth-century experiments in the hunting and gathering lifestyle. And as such they can tell us important things about patterns in prehistory even if the groups studied are not pristine remnants of that prehistory. For example, they can presumably tell us about a people's ability to extract balanced diets from wild resources with simple technology lacking any source of energy other than human power. They can tell us about the relative efficiency of different foraging techniques and extraction methods connected with hunting big game, hunting smaller animals, fishing, shellfishing, and gathering and processing various vegetable foods. They can also tell us something about the effect of small group size and mobility on the transmission of infectious disease.

A broader collection of comparative data on twentieth-century hunter-gatherer nutrition from around the world (Cohen 1989) suggests that contemporary hunter-gatherers (with the exception of those in the Arctic, where vegetable foods are scarce) seem routinely to enjoy relatively eclectic and thus well-balanced diets of fresh foods. Moreover, their typical practice of exploiting a relatively wide range of soils tends to minimize the impact of specific nutrient deficiencies (such as iodine) that are associated with particular soils. As a result, these groups are, for the most part, well nourished at least by contemporary developing-world standards; and they are conspicuously well nourished in comparison to the modern world's poor.

Where contemporary hunter-gatherers coexist with farming populations, as is now commonly the case, hunter-gatherers typically operate as specialists who trade protein, vitamins, and variety foods to farmers in exchange for calories (Williams 1974; Peterson

1978; Griffin 1984). It would appear, therefore, that hunting and gathering diets are almost always relatively nutritious in terms of variety and quality but potentially lacking in calories.

Nonetheless, caloric intake by hunter-gatherers appears sufficient when compared to modern developing-world populations. For example, San caloric intake, although considered marginal, is estimated at 2,000 to 2,100 kcal per person per day, although it falls somewhat below 2,000 kcal per person per day in poor seasons (Lee 1969; Wilmsen 1978; Tanaka 1980). Yet this compares favorably with estimated caloric intake in developing-world countries such as India and China, which *averages* only 1,800 to 2,200 kcal (Bunting 1970; Clark and Haswell 1970; Pellet 1983). Moreover, it compares very favorably with estimates for modern urban poor, who may take in as few as 1,100 to 1,500 kcal per person per day (Basta 1977).

Contemporary hunter-gatherers also receive a relatively large part of their diet from animal products. This is in the range of 20 to 40 percent of the diet, which is about the same as that estimated for affluent modern Western people but well above the modern world average. Daily animal protein intake among the San, for example, is estimated by various sources at approximately 30 to 50 grams per person per day (Lee 1968; Wilmsen 1978; Tanaka 1980; Silberbauer 1981), which far exceeds an estimated average of 7 to 10 grams of animal protein per person in modern developing-world countries and among the world's poor (Basta 1977; Peterson 1978).

It should also be noted, in response to the observation that contemporary hunter-gatherers may have low caloric intake, that they live in some of the world's poorest environments and consequently those most difficult to exploit for food. In fact, judging from the nutritional experience of other contemporary hunter-gatherers, it would appear that the !Kung San, although typical in the variety of their diets, are actually somewhat below hunter-gatherer average in their caloric and protein intake (Hawkes and O'Connell 1985; O'Connell, Hawkes, and Blurton-Jones 1988; Cohen 1989). Populations such as the Hadza of Tanzania, who live in a richer foraging area, are estimated to get 3,000 kcal and 50 to 250 grams of meat protein per person per day (O'Connell et al. 1988). Indeed, groups like the Hadza appear to be a better model for prehistory than the San because they live in the same kinds of environments as early human beings. Yet even the Hadza frequent an area now partly depleted of big game (Hawkes, O'Connell, and Blurton-Jones 1992).

Infection and Nutrition

Another important but not always apparent factor that must be considered in assessing the diets of hunter-gatherers is their comparative freedom from parasites, which affect the nutritional value of diets in a variety of ways (Scrimshaw, Taylor, and Gordon

1968; Beisel 1982). Parasites can cause diarrhea, speeding up the flow of nutrients through the intestine, and therefore interfere with nutrient absorption from the intestine into the bloodstream. In some diseases, such as malaria or hookworm, the parasites destroy or consume human tissues (in these cases, red blood cells), which must be replaced. Other parasites, such as tapeworms, simply compete with humans for the vitamins and minerals in our digestive tract. And infection may actually cause the body to deny itself nutrients as a means of destroying the invader, as can be the case with the body withholding iron (Weinberg 1974, 1992).

Parasite load is, to a large degree, a function of habitat. Warmth and moisture generally encourage the survival and transmission of parasites, so that tropical forest hunter-gatherers such as the Pygmies of Zaire have higher parasite loads than those in drier or colder climates (Price et al. 1963; cf. Heinz 1961).

But the parasite load also tends to increase with the size and density of the human population and with permanence of human settlement, regardless of climate, since larger accumulations of filth, of people, and of stored food all facilitate parasite transmission. Diarrhea-causing organisms, for example, are typically transmitted by fecal-oral infection, in which feces contaminate food and water supplies, a problem that is relatively rare among small and mobile groups. Hookworm infection also thrives on human crowding. The worms grow from eggs deposited on the ground in human feces. They then penetrate human skin (usually the soles of feet) and find their way "back" to the intestine, where they live by consuming red blood cells while shedding a new generation of eggs. Obviously, people on the move are less likely to contaminate the soil around them.

Tapeworms, whose life cycles commonly include both people and domestic animals, are also rare in societies that keep no animals but obtain their meat by hunting wild game. Tapeworms typically are passed to domestic animals such as cows and pigs by human feces. The proximity of domestic animals as well as the density of both human and animal populations facilitates transmission.

The !Kung San avoid most such parasites (Heinz 1961). They do suffer from hookworm because even their desert habitat, mobile habits, and small groups do not entirely prohibit transmission; but they suffer only a fairly mild infestation that is not generally sufficient to promote anemia, the main danger of hookworm (see Truswell and Hansen 1976).

In short, increased parasite load diminishes the quality of nutrition, but hunter-gatherers suffer less of a nutritional loss to parasites than other societies in the same environments. The consequence is that hunter-gatherers require smaller dietary intakes than people in those other societies.

Hunting and Gathering Populations of Prehistory

Reason to believe that the hunting and gathering populations of prehistory were at least as well nourished as their contemporary counterparts can be gleaned from comparing the environments in which prehistoric hunter-gatherers chose to live and those to which their modern counterparts are confined by the pressures of competition with more powerful neighbors. Prehistoric, but biologically modern, human hunter-gatherers seem to have expanded first through relatively game-rich environments, savannas, steppes, and open forests. Occupation of the deserts and jungles in which most hunting and gathering groups now find themselves is a relatively recent phenomenon.

Hunters also seem initially to have focused on medium- to large-sized animal prey plus a relatively narrow spectrum of plant foods. Small game, seeds to grind, fish and shellfish (at least consistently and in quantity) all appear to be relatively recent additions to the human larder. The additions were made in the Mesolithic period of prehistory (Cohen 1977) and were associated with what K. V. Flannery (1973) has called the "broad spectrum revolution," which took place within, approximately, the last 15,000 years. The use of secondary habitats and the adoption of demonstrably inefficient foraging techniques suggest strongly that the diets of hunter-gatherers began to decline under the pressure of their own populations almost from the time that the efficient hunter, *Homo sapiens*, first emerged to spread out around the world.

Actual tests of the relative efficiency of various foraging techniques indicate that prehistoric hunter-gatherers in environments richer in large game than that occupied by contemporary counterparts would have fared well in comparison to contemporary groups. Indeed, numerous investigations point out that when available, large game animals can be taken and converted to food far more efficiently than most other wild resources. Data provided by Stuart Marks (1976) and recalculated by the author (Cohen 1989) suggest, for example, that big game hunters without modern rifles or shotguns in game-rich environments may obtain an *average* of as much as 7,500 to 15,000 kcal for every hour of hunting. Many other studies also suggest that large game, although relatively scarce in the modern world, can be taken with great efficiency once encountered, even if hunters do not use modern firearms (Jones 1980; Blackburn 1982; Rowley Conway 1984; Hawkes and O'Connell 1985).

By contrast, most of the resources taken by contemporary hunters, including small game, fish, shellfish, and small-seeded vegetables, are far less efficient to gather and convert to food. Estimates of shellfish-gathering efficiency, for example, suggest that it produces about 1,000 kcal per hour of work; hunting small game may average no more than 500 to 800

kcal per hour; small seed processing also produces only about 500 to 1,000 kcal per hour (Jones 1980; Winterhalder and Smith 1981; Rowley Conway 1984; Cohen 1989). Collection of nuts may, however, constitute a partial exception. Brazil nuts, for example, can be harvested at rates that provide kcals comparable to hunting. But the nuts must still be cracked and processed into food, both relatively time-consuming activities. Similarly, anadromous (migratory) fish can be harvested very efficiently but only after large weirs have been constructed (Werner 1983; Rowley Conway 1984).

Interestingly, the relative efficiency of hunting large game (when available) appears to hold whether foragers use iron or stone tools. In fact, metal tools apparently add relatively little to hunting efficiency, although they add significantly to the efficiency of gathering vegetable foods, not to mention growing them. In a Stone Age world, therefore, the advantage of being a big game hunter would have been substantially greater than even these modern comparative tests of various economic activities undertaken with metal tools suggest (Colchester 1984; Harris 1988). Hunting large game with spears is also clearly more efficient than hunting smaller game with bows and arrows and small projectile points or nets or probably even primitive muskets – the point being that “improvements” in hunting technology did not offset the loss of efficiency that occurred as prey size became smaller. In addition, hunting large game would have clearly been more efficient than harvesting wild wheat or farming wheat with stone tools such as the sickles and grindstones that appear in human tool kits relatively late in prehistory (Russell 1988). In short, a decline in available big game was apparently more important than any technological innovation in affecting foraging choices and determining the overall efficiency of the economy. The ultimate adoption of farming seems to have been only one in a long series of strategies adopted to offset diminishing returns.

One further point is worth making. Modern hunter-gatherers, such as the !Kung or even the Hadza, who contend with game-depleted environments or those with legal hunting restrictions must clearly be less efficient in putting food on the table than their (and our) prehistoric forebears.

The data also indicate that prehistoric hunters in game-rich environments would have consumed diets containing a larger proportion of meat than most of their contemporary counterparts. Yet as John Speth (1988) has argued, there are limits to the proportion of meat in the diet that a human being can tolerate since meat, without commensurate carbohydrates or fats, is calorically expensive to process. Moreover, meat is a diuretic, so that people like the San, who live in hot deserts where water and calories are both scarcer than protein, may limit meat consumption to conserve water.

The Evidence of Prehistoric Skeletons

There is also a good deal of direct evidence (most of it gathered since 1960) to support the hypothesis that prehistoric hunter-gatherers were relatively well nourished. Their skeletons, often in large numbers, have been analyzed from various regions of the world. In more than 20 areas of the globe (but mostly in North America) it is possible to use these remains to make comparative analyses of the nutrition and health of two or more prehistoric populations representing different stages in the evolution of food technology (Cohen and Armelagos 1984). For example, in specific cases we can compare hunter-gatherers to the farmers who succeeded them in the same region; or compare early hunters to later foragers; or incipient farmers to intensive farmers, and so forth.

Such analyses generally confirm that infection and associated malnutrition become more common as small groups become larger and more sedentary. The skeleton displays nonspecific infections called periostitis when only the outer surface of the bone is affected and osteomyelitis when the infection penetrates deep into the medullary cavity of the bone. Osteomyelitis is rarely found in prehistoric skeletons, but periostitis is routinely found to have been more common in larger and more sedentary groups and can probably be taken as an index of the prevalence of other infectious diseases. In addition, other types of infection can occasionally be glimpsed. For example, a comparison of mummified populations from Peru (Allison 1984) demonstrates an increase in intestinal parasites with sedentism. A comparison of preserved fecal material from different archaeological layers in the American Southwest also demonstrates an increase in parasites with the adoption of sedentism (Reinhard 1988).

Other such evidence can be found in the characteristic lesions on the skeleton left by diseases such as yaws, syphilis, leprosy, and tuberculosis, all of which increase with density or appear only in relatively civilized populations. Tuberculosis appears to be almost entirely a disease of relatively recent, civilized populations in both the Old World and the New (Buikstra 1981; Cohen and Armelagos 1984). Yaws (a nonvenereal disease caused by a spirochete identical to the one that causes syphilis) has been shown to increase with population density among New World Indians (Cohen and Armelagos 1984).

Skeletons also provide fairly specific signs of anemia, or lack of sufficient red blood cell function. The condition is called porotic hyperostosis and cribra orbitalia and appears as a thickening and porosity of the bones of the cranium and eye orbits in response to the enlargement of marrow cavities where red blood cells are formed.

Anemia can result from inadequate dietary intake of iron associated with diets high in maize and other cereals, since the cereals are poor sources of iron and

may actually interfere with iron absorption. However, increasingly, anemia is thought to reflect the secondary loss of iron to parasites such as hookworm, and losses in fighting diseases such as tuberculosis, and even the body's own sequestering of iron to fight infection (Weinberg 1974, 1992; Stuart-Macadam 1992). In one particular archaeological sequence from the American Southwest, in which preserved human feces have been examined, anemia was shown to relate to the frequency of parasitic worms in stools rather than to diet (Reinhard 1988, 1992). But whatever the cause, anemia seems to have been primarily a disease of more civilized or sedentary farmers rather than hunter-gatherers everywhere they have been studied, and it *increases* through time in association with group size and sedentism in almost all reported archaeological sequences (Cohen and Armelagos 1984).

One other dietary deficiency disease, rickets in children and osteomalacia in adults, can be diagnosed in the skeleton. Soft or malformed bones resulting from improper calcification can result from lack of calcium or lack of vitamin D in the diet. Most commonly, however, it occurs from lack of exposure to sunlight, because most vitamin D is produced in the skin as the result of exposure to ultraviolet radiation. The archaeological record suggests that rickets is very rare among prehistoric hunter-gatherers but common, as one might predict, among the inhabitants of smog-bound urban ghettos in the last few centuries (Steinbock 1976; Cohen and Armelagos 1984).

Changes in human growth and stature may also reflect a decline in the quality of human nutrition through time. Many authorities consider average stature to be a fairly reliable indicator of nutritional status (see Fogel et al. 1983), and certainly the increase in European and American stature in the last century has been viewed as evidence of improving nutrition. But for centuries prior to the nineteenth century, decline was the predominant trend in human stature.

The first biologically modern human populations of hunter-gatherers throughout Europe and areas of Asia including India seem to have been relatively tall. Unquestionably these Paleolithic hunters were taller than the Mesolithic foragers and Neolithic farmers that came after them (Angel 1984; Kennedy 1984; Meiklejohn et al. 1984; Smith, Bar-Yosef, and Sillen 1984), and the populations of eighteenth-century Europe to which we compare ourselves with considerable pride were among the shortest human groups that ever lived (Fogel 1984).

Retarded growth may also be identified in the skeletons of children whose bones suggest that they were smaller for their age (as determined by the state of tooth formation and eruption at death) than children living at some other time or place. For example, skeletons of children from the Dickson Mounds archaeological site in Illinois suggest that childhood growth was retarded in a farming population when

compared to that of their foraging forebears (Goodman et al. 1984). In addition, malnutrition may show up as premature osteoporosis, the thinning of the outer, solid, cortical portions of bones. This condition seems to be more important in farmers or later populations than in prehistoric hunter-gatherers (e.g., Stout 1978; Smith et al. 1984).

Finally, the adult human skeleton displays scars of biological or nutritional stresses felt in childhood, particularly those associated with weanling malnutrition and weanling diarrhea. Illness while teeth are growing can result in irregularities in tooth enamel that leave a permanent record of stress in the form of visible lines called enamel hypoplasia or microscopic defects called Wilson bands (see Rose, Condon, and Goodman 1985). Prehistoric hunter-gatherers fairly typically show lower rates of these defects than do the farming and civilized populations that followed them, confirming the observation that hunter-gatherer children endured significantly less weanling stress than did farmers or other more "civilized" neighboring populations (Cohen and Armelagos 1984; Cohen 1989).

It is true that some critics object to such conclusions by observing that the use of skeletal indicators of stress in prehistoric populations may be misleading – that, for various reasons and in various ways, skeletons may provide an unrepresentative or biased sample of a once-living population. (For details of the argument see Wood et al. 1992, and Cohen forthcoming.) Yet skeletal evidence accords well with ethnographic observations and with predictions of epidemiology. In other words, infection not only increases with sedentism in skeletal populations but also increases in many ethnographic or historically described skeleton groups. Moreover, as already discussed, contemporary hunter-gatherers (and not just prehistoric hunter-gatherers) seem to be well protected against anemia. Put plainly, skeletal data when checked against other results appear to be giving us an accurate and coherent picture of past health and nutrition (Cohen 1989, 1992).

The Texture of the Diet

Although controversies remain about the quality and quantity of food available to both modern and ancient hunter-gatherers, there is little dispute that there have been significant changes in dietary texture throughout history. Contemporary hunter-gatherers as a group (and, presumably, their prehistoric counterparts) eat foods that differ in texture from modern diets in three important ways: Wild foods are comparatively tough to chew; they are high in bulk or fiber; and, with the occasional exception of honey, they lack the high concentrations of calories found in many modern processed foods. These textural differences have several effects on human development and health. First, individuals raised on hunter-gatherer diets develop different occlusion of their teeth in which the upper

and lower incisors meet edge to edge. The modern "normal" pattern of slight overbite is actually a consequence of modern soft diets (Brace 1986). Hunter-gatherer diets also generate significantly more tooth wear than do modern diets, so that in contrast to civilized populations, hunter-gatherers are at risk of literally wearing out their teeth.

However, modern diets rich in sweet and sticky substances are far more cariogenic. Significant tooth decay is, for the most part, a relatively recent phenomenon. Historically, rates of caries increased dramatically with the adoption of pottery, grindstones, and farming (which made softer diets possible) and again with the production of refined foods in the last few centuries (Powell 1985). In fact, the difference in caries rates between ancient hunter-gatherers and farmers is so pronounced that many archaeologists use the rate of caries in archaeological skeletons to help distinguish between prehistoric hunters and farmers (see Turner 1979; Rose et al. 1984).

Coarse textured foods have a particularly important effect on two segments of the population, the very old with badly worn teeth and the very young. The problem of feeding the very young may require a mother to delay weaning, and this may help to explain the relatively low fertility of the !Kung – a phenomenon that some, but not all, modern hunter-gatherer populations seem to experience one yet that may have been the experience of hunter-gatherers in prehistory (Cohen 1989; cf. Wood 1990). Without soft foods to wean their children and with "baby food" very difficult to prepare, hunter-gatherers may be forced to nurse both intensively and for a relatively long period. Lactation, especially in combination with low caloric intake and high energy output, is known to exert contraceptive effects (Konner and Worthman 1980; Habicht et al. 1985; Ellison 1990).

But the adoption of cereals and grindstones to prepare gruel by Mesolithic and Neolithic populations would have simplified the problem of feeding the very young whether or not it improved nutrition. And early weaning in turn would help to explain an apparent increase in the growth rate of the human population after the adoption of farming in the Neolithic period, even though no corresponding improvement in health or longevity can be documented (Cohen and Armelagos 1984; Cohen 1989).

The lack of refined foods available to them may also explain the relative immunity of hunting and gathering populations (and many other populations) to diet-related diseases that plague twentieth-century Western populations. For example, the relatively low calorie-for-volume content of hunter-gatherer diets helps to explain the relative scarcity of obesity and obesity-related conditions among such groups. (Even wild animals that must work for a living are relatively lean in comparison to their modern domestic counterparts.) Adult-onset diabetes is very rare in "primitive" societies, although studies in various parts of the world suggest

that the same individuals may be diabetes-prone when switched to Western diets (Neel 1962; Cohen 1989). Similarly, high blood pressure is essentially unknown among hunter-gatherer groups who enjoy low sodium (and perhaps also high potassium or calcium) diets, although the same groups develop high blood pressure when "civilized" (Cohen 1989).

High-fiber diets among hunter-gatherers and other "primitive" groups also affect bowel transit time. Members of such groups typically defecate significantly more often than "civilized" people. In consequence, diseases associated with constipation such as appendicitis, diverticulosis, varicose veins, and bowel cancer are all relatively rare among hunter-gatherers (and non-Western populations in general) and are thought to result at least in part from modern, Western low-bulk diets (Burkitt 1982).

In summary, a number of lines of evidence from archaeology, from prehistoric skeletons, and from the study of contemporary populations indicate that small, mobile human groups living on wild foods enjoy relatively well-balanced diets and relatively good health. Indeed, the available evidence suggests that hunter-gatherer diets remain well balanced even when they are low in calories. The data also show that per capita intake of calories and of protein has declined rather than increased in human history for all but the privileged classes. The predominant direction of prehistoric and historic change in human stature has been a decline in size despite the "secular trend" among some Western populations of the last century. Prehistoric remains of more sedentary and larger groups commonly display an increase in general infection and in specific diseases (such as yaws and tuberculosis), combined with an increase in porotic hyperostosis (anemia) and other signs of malnutrition.

Mark Nathan Cohen

Bibliography

- Allison, M. J. 1984. Paleopathology in Peruvian and Chilean mummies. In *Paleopathology at the origins of agriculture*, ed. M. N. Cohen and G. J. Armelagos, 515–30. New York.
- Angel, J. L. 1984. Health as a crucial factor in changes from hunting to developed farming in the eastern Mediterranean. In *Paleopathology at the origins of agriculture*, ed. M. N. Cohen and G. J. Armelagos, 51–74. New York.
- Baker, Brenda, and George J. Armelagos. 1988. The origin and antiquity of syphilis. *Current Anthropology* 29: 703–20.
- Basta, S. S. 1977. Nutrition and health in low income urban areas of the Third World. *Ecology of food and nutrition* 6: 113–24.
- Beisel, W. R. 1982. Synergisms and antagonisms of parasitic diseases and malnutrition. *Review of Infectious Diseases* 4: 746–55.
- Binford, L. R. 1968. Post Pleistocene adaptations. In *New perspectives in archaeology*, ed. S. R. Binford and L. R. Binford, 313–36. Chicago.

- Blackburn, R. H. 1982. In the land of milk and honey: Okiek adaptation to their forests and neighbors. In *Politics and history in band societies*, ed. E. Leacock and R. B. Lee, 283-306. Cambridge.
- Blurton-Jones, N., and P. M. Sibley. 1978. Testing adaptiveness of culturally determined behavior: Do the Bushmen women maximize their reproductive success? In *Human behavior and adaptation*. Society for the Study of Human Biology, Symposium 18, 135-57. London.
- Boserup, Ester. 1965. *The conditions of agricultural growth*. Chicago.
- Brace, C. L. 1986. Eggs on the face. . . . *American Anthropologist* 88: 695-7.
- Bronte-Stewart, B., O. E. Budtz-Olsen, J. M. Hickey, and J. F. Brock. 1960. The health and nutritional status of the !Kung Bushmen of South West Africa. *South African Journal of Laboratory and Clinical Medicine* 6: 188-216.
- Buikstra, J., ed. 1981. *Prehistoric tuberculosis in the Americas*. Evanston, Ill.
- Bunting, A. H. 1970. *Change in agriculture*. London.
- Burkitt, Denis P. 1982. Dietary fiber as a protection against disease. In *Adverse effects of foods*, ed. E. F. Jelliffe and D. B. Jelliffe, 483-96. New York.
- Clark, C., and M. Haswell. 1970. *The economics of subsistence agriculture*. Fourth edition. London.
- Cohen, Mark N. 1977. *The food crisis in prehistory*. New Haven, Conn.
1989. *Health and the rise of civilization*. New Haven, Conn.
1992. Comment. In The osteological paradox: Problems of inferring prehistoric health from skeletal samples. *Current Anthropology* 33: 358-9.
- forthcoming. The osteological paradox - reconsidered. *Current Anthropology*.
- Cohen, Mark N., and G. J. Armelagos, eds. 1984. *Paleopathology at the origins of agriculture*. New York.
- Colchester, Marcus. 1984. Rethinking stone age economics: Some speculations concerning the pre-Columbian Yanomama economy. *Human Ecology* 12: 291-314.
- Draper, H. H. 1977. The aboriginal Eskimo diet. *American Anthropologist* 79: 309-16.
- Ellison, P. 1990. Human ovarian function and reproductive ecology: New hypotheses. *American Anthropologist* 92: 933-52.
- Flannery, K. V. 1969. Origins and ecological effects of early domestication in Iran and the Near East. In *The domestication and exploitation of plants and animals*, ed. P. J. Ucko and J. W. Dingleby, 73-100. London.
1973. The origins of agriculture. *Annual Reviews in Anthropology* 2: 271-310.
- Fogel, R. W. 1984. *Nutrition and the decline in mortality since 1700: Some preliminary findings*. National Bureau of Economic Research, Working Paper 1402. Cambridge, Mass.
- Fogel, R. W., S. L. Engerman, R. Floud, et al. 1983. Secular changes in American and British stature and nutrition. In *Hunger and history*, ed. R. I. Rotberg and T. K. Rabb, 247-83. Cambridge and New York.
- Goodman, A., J. Lallo, G. J. Armelagos, and J. C. Rose. 1984. Health changes at Dickson Mounds, Illinois, A.D. 950-1300. In *Paleopathology at the origins of agriculture*, ed. M. N. Cohen and G. J. Armelagos, 271-306. New York.
- Griffin, P. B. 1984. Forager resource and land use in the humid tropics: The Agta of northeastern Luzon, the Philippines. In *Past and present in hunter-gatherer studies*, ed. Carmel Schrire, 99-122. New York.
- Habicht, J-P, J. Davanzo, W. P. Buttz, and L. Meyers. 1985. The contraceptive role of breast feeding. *Population Studies* 39: 213-32.
- Harris, Marvin. 1988. *Culture, people, nature*. Fifth edition. New York.
- Hawkes, Kristen, and J. F. O'Connell. 1985. Optimal foraging models and the case of the !Kung. *American Anthropologist* 87: 401-5.
- Hawkes, Kristen, J. F. O'Connell, and N. Blurton-Jones. 1992. Hunting income patterns among the Hadza: Big game, common goods, foraging goals, and the evolution of human diet. In *Foraging strategies and natural diet of monkeys, apes, and humans*, ed. A. Whiten and E. M. Widdowson, 83-91. New York.
- Heinz, H. J. 1961. Factors governing the survival of Bushmen worm parasites in the Kalahari. *South African Journal of Science* 8: 207-13.
- Hobbes, T. 1950. *Leviathan*. New York.
- Jelliffe, D. B., J. Woodburn, F. J. Bennett, and E. P. F. Jelliffe. 1962. The children of the Hadza hunters. *Tropical Paediatrics* 60: 907-13.
- Jones, Rhys. 1980. Hunters in the Australian coastal savanna. In *Human ecology in savanna environments*, ed. D. Harris, 107-47. New York.
- Kennedy, K. A. R. 1984. Growth, nutrition and pathology in changing paleodemographic settings in South Asia. In *Paleopathology at the origins of agriculture*, ed. M. N. Cohen and G. J. Armelagos, 169-92. New York.
- Konner, M., and C. Worthman. 1980. Nursing frequency, gonad function and birth spacing among !Kung hunter gatherers. *Science* 207: 788-91.
- Lee, R. B. 1968. What hunters do for a living or how to make out on scarce resources. In *Man the Hunter*, ed. R. B. Lee and I. DeVore, 30-48. Chicago.
1969. !Kung Bushman subsistence: An input-output analysis. In *Ecological studies in cultural anthropology*, ed. A. P. Vayda, 47-79. Garden City, N.Y.
- Lee, R. B., and I. DeVore, eds. 1968. *Man the hunter*. Chicago.
- Mann, G. V., O. A. Roels, D. L. Price, and J. M. Merrill. 1963. Cardiovascular disease in African Pygmies. *Journal of Chronic Diseases* 14: 341-71.
- Marks, Stuart. 1976. *Large mammals and a brave people*. Seattle.
- Meiklejohn, C., Catherine Schentag, Alexandra Venema, and Patrick Key. 1984. Socioeconomic changes and patterns of pathology and variation in the Mesolithic and Neolithic of Western Europe. In *Paleopathology at the origins of agriculture*, ed. M. N. Cohen and G. J. Armelagos, 75-100. New York.
- Metz, J. D., D. Hart, and H. C. Harpending. 1971. Iron, folate and vitamin B₁₂ nutrition in a hunter-gatherer people: A study of !Kung Bushmen. *American Journal of Clinical Nutrition* 24: 229-42.
- Neel, J. V. 1962. Diabetes mellitus: A "thrifty" genotype rendered detrimental by progress? *American Journal of Human Genetics* 14: 355-62.
- O'Connell, J. F., K. Hawkes, and N. Blurton-Jones. 1988. Hadza scavenging: Implications for Plio/Pleistocene hominid subsistence. *Current Anthropology* 29: 356-63.
- Pellet, P. 1983. Commentary: Changing concepts of world malnutrition. *Ecology of Food and Nutrition* 13: 115-25.
- Peterson, J. T. 1978. Hunter-gatherer farmer exchange. *American Anthropologist* 80: 335 - 51.
- Powell, M. 1985. The analysis of dental wear and caries for dietary reconstruction. In *The analysis of prehistoric diets*, ed. R. I. Gilbert and J. H. Mielke, 307-38. New York.

- Price, D. L., G. V. Mann, O. A. Roels, and J. M. Merrill. 1963. Parasitism in Congo Pygmies. *American Journal of Tropical Medicine and Hygiene* 12: 83-7.
- Reinhard, Carl. 1988. Cultural ecology of prehistoric parasites on the Colorado Plateau as evidenced by coprology. *American Journal of Physical Anthropology* 77: 355-66.
1992. Patterns of diet, parasites and anemia in prehistoric Western North America. In *Diet, demography and disease*, ed. P. S. Macadam and S. Kent, 219-38. Chicago.
- Rose, Jerome C., B. A. Burnett, M. S. Nassaney, and M. W. Blauer. 1984. Paleopathology and the origins of maize agriculture in the lower Mississippi Valley and Caddoan culture areas. In *Paleopathology at the origins of agriculture*, ed. M. N. Cohen and G. J. Armelagos, 393-424. New York.
- Rose, Jerome C., K. W. Condon, and A. H. Goodman. 1985. Diet and dentition: Developmental disturbances. In *The analysis of prehistoric diets*, ed. R. I. Gilbert and J. Mielke, 281-306. New York.
- Rowley Conway, P. 1984. The laziness of the short distance hunter: The origins of agriculture in western Denmark. *Journal of Anthropological Archaeology* 38: 300-24.
- Russell, Kenneth W. 1988. After Eden. *British Archaeological Reports, International Series 391*. Oxford.
- Sahlins, M. 1968. Notes on the original affluent society. In *Man the Hunter*, ed. R. B. Lee and I. DeVore, 85-8. Chicago.
- Schrire, Carmel. 1980. An inquiry into the evolutionary status and apparent history of the San hunter-gatherers. *Human Ecology* 8: 9-32.
- ed. 1984. *Past and present in hunter-gatherer studies*. New York.
- Scrimshaw, N., C. Taylor, and J. Gordon. 1968. *Interactions of nutrition and infection*. Geneva.
- Silberbauer, George B. 1981. *Hunter and habitat in the central Kalabari*. Cambridge.
- Smith, Patricia, O. Bar-Yosef, and A. Sillen. 1984. Archaeological and skeletal evidence of dietary change during the late Pleistocene/early Holocene in the Levant. In *Paleopathology at the origins of agriculture*, ed. M. N. Cohen and G. J. Armelagos, 101-36. New York.
- Speth, John. 1988. *Hunter-gatherer diet, resource stress and the origins of agriculture*. Symposium on Population Growth, Disease and the Origins of Agriculture. New Brunswick, N.J. Rutgers University.
- Steinbock, R. T. 1976. *Paleopathological diagnosis and interpretation*. Springfield, Ill.
- Stout, S. D. 1978. Histological structure and its preservation in ancient bone. *Current Anthropology* 19: 600-4.
- Stuart-Macadam, P. 1992. Anemia in past human populations. In *Diet, demography and disease*, ed. P. S. Macadam and S. Kent, 151-76. Chicago.
- Stuart-Macadam, P., and S. Kent, eds. 1992. *Diet, demography and disease*. Chicago.
- Tanaka, J. 1980. *The San: Hunter gatherers of the Kalabari*. Tokyo.
- Truswell, A. S., and J. D. L. Hansen. 1976. Medical Research among the !Kung. In *Kalabari Hunter Gatherers*, ed. R. B. Lee and I. DeVore, 166-95. Cambridge.
- Turner, Christie. 1979. Dental anthropological indicators of agriculture among the Jomon People of central Japan. *American Journal of Physical Anthropology* 51: 619-35.
- Wehmeyer, A. S., R. B. Lee, and M. Whiting. 1969. The nutrient composition and dietary importance of some vegetable foods eaten by the !Kung Bushmen. *Tydkrif vir Geneeskunde* 95: 1529-30.
- Weinberg, E. D. 1974. Iron and susceptibility to infectious disease. *Science* 184: 952-6.
1992. Iron withholding in prevention of disease. In *Diet, demography and disease*, ed. P. Stuart-Macadam and S. Kent, 105-50. Chicago.
- Werner, Dennis. 1983. Why do the Mekranoti trek? In *The adaptive responses of Native Americans*, ed. R. B. Hames and W. T. Vickers, 225-38. New York.
- Williams, B. J. 1974. *A model of band society*. Memoirs of the Society for American Archaeology No. 39. New York.
- Wilmsen, Edwin. 1978. Seasonal effects of dietary intake on the Kalahari San. *Federation Proceedings* 37: 65-72.
- Winterhalder, Bruce, and E. A. Smith, eds. 1981. *Hunter gatherer foraging strategies*. Chicago.
- Wood, James. 1990. Fertility in anthropological populations. *Annual Review of Anthropology* 19: 211-42.
- Wood, James, G. R. Milner, H. C. Harpending, and K. M. Weiss. 1992. The osteological paradox: Problems of inferring prehistoric health from skeletal samples. *Current Anthropology* 33: 343-70.
- Woodburn, James. 1968. An introduction to Hadza ecology. In *Man the Hunter*, ed. R. B. Lee and I. DeVore, 49-55. Chicago.



PART II

Staple Foods: Domesticated Plants and Animals

Part II, with its almost 60 chapters that concentrate on staple foods (most of the fruits are treated in Part VIII), constitutes the largest portion of this work. Yet the space devoted to it seems more than justified in light of the immensity of the effort that humans have invested in domesticating the plants of the fields and the animals of the barnyard. In the case of plants, the effort began with the harvesting of wild grains and roots, which probably became a seasonal activity for some hunter-gatherers as the last Ice Age receded and the large mammals – mammoths, mastodons, giant sloths, and the like – were embarking on their journey to extinction. The next leap – a giant one – was from locating wild grains for harvest to planting them in permanent places and then tinkering with them and the soil so they would do a better job of growing.

Wolves were domesticated to become dogs toward the end of the Paleolithic. Like their human companions, they were hunters. But with the beginning of farming, a host of other animals followed them into domestication. In the Old World, goats and sheep, which fed on wild grasses but had no objection to domesticated ones, were probably initially perceived by early farmers as competitors – if not outright thieves – to be fenced out or chased away. But with the precedent of dog domestication to guide them, people began to capture these ruminants and eventually raised them in captivity. In many cases, the purpose seems to have been a supply of animals close at hand for ceremonial sacrifice. But even if meat, milk, hides, hair, and wool were secondary products at first, they were, nonetheless, important ones that quickly rose to primacy. Pigs, cattle, water buffalo, horses, even camels and, later on, chickens may have also been initially sought for sacrifice and then, after

domestication, exploited for their other products, including labor and transportation.

Such a range of large animals that could be domesticated was, however, restricted to the core of the Old World. On its periphery, in Africa south of the Sahara, trypanosomal infection delivered by the tsetse fly often discouraged livestock keeping, and, in the New World, save for dogs, turkeys, and llamas and llama-like creatures, there was relatively little animal domestication carried out. This meant for many an absence of animal fats in the diet, which nature remedied, to some extent, by making the fat in avocados available to many Americans, and palm oil and coconut palms to most in the tropical world.

In the Americas, a shortage of animal protein also meant a heavy reliance on maize, beans, squashes, and potatoes, along with manioc in tropical and subtropical regions; in Africa, yams, millet, sorghum, and a kind of rice sustained life; and in Oceania, taro served as an important staple, along with the sweet potato – an American plant that somehow had diffused throughout that vast area long before the Europeans arrived.

This mystery, along with another occasioned by Old World onions and garlic noted by the expedition of Hernando Cortes in Mexico, may never be solved, and many other such mysteries probably never had a chance to come to light because of the process of food globalization set in motion after 1492. Animals, grains, and vegetables domesticated in the Old World flourished in the New. Beef, pork, and cheese, for example, fitted so naturally into Mexican cuisine that it is difficult to appreciate that they have not been there forever. In similar fashion, the American plants revolutionized the cuisines of the other lands of the globe. Potatoes spread from Ireland to Russia, fields of

maize sprang up as far away as China, and manioc combined with maize to cause a population explosion in Africa. Chilli (we stand by our expert's spelling) peppers raced around the globe to add fire to African soups and Indian curries; the tomato was married to pasta in Italy; and, in the other direction, Africa sent varieties of field peas to enliven regional dishes of the Americas and okra to make gumbos.

The continuation of food globalization since 1492 – especially since the 1960s – has caused increasing concern about the atrophy of regional cuisines on the one hand, and the spread of “fast food” on the other. But, as a 1992 Smithsonian exhibition made clear, the “Seeds of Change” discussed in Part II have been broadcast with increasing intensity around the world since the Columbian voyages, and not always with good effect.



II.A

Grains

II.A.1. Amaranth

A robust annual herb with seeds as small as mustard seeds, amaranth belongs to the genus *Amaranthus* of the family Amaranthaceae, with 50 to 60 species scattered throughout the world in wild and domesticated forms. Most are weeds, such as pigweed (*A. retroflexus*), which commonly invades gardens in the United States, whereas others are raised as ornamentals. The genus derives its name from the Greek meaning “unfading,” “immortal,” or “not withering” because the flowers remain the same after they are dried. Poets have favored the amaranth, therefore, as a symbol of immortality (Sauer 1976; Berberich 1980; Tucker 1986; Amaranthus 1991).

It is as a food, however, that amaranth was and is most important to human populations, either for its leaves or seeds. Two species of amaranth, *A. tricolor* and *A. dubius*, are popular among Chinese-Americans for soup and salad greens. But the most versatile and nutritious are the grain amaranths, because the genus *Amaranthus*, although a non-grass, is capable of producing great amounts of edible grain (Cramer 1987). Three species of amaranth that were domesticated in the Americas and are commonly utilized as grain are *A. hypochondriacus*, known as “prince’s feather” in England, from northwestern and central Mexico; *A. cruentus* of southern Mexico and Central America, whose greens are widely utilized in Africa; and *A. caudatus* of the Andes, known as “love-lies-bleeding” in the United States (Sauer 1976; Cole 1979). The first two species are now cultivated in the United States as seed grains, but *A. caudatus* does not do as well in the United States as in the Andes (Tucker 1986).

The amaranth plant grows from 1 to 10 feet tall in an erect or spreading form. It is a broad-leaved plant that bears up to 500,000 black, red, or white seeds on a single large seedhead, made up of thick fingerlike spikes (Cole 1979; Tucker 1986). The leaves are “often

variegated with bronze, red, yellow, or purple blotches” (Wister 1985), and the flowers may be orange, red, gold, or purple. Its beautiful colors have led people throughout the world to raise amaranth species as ornamental plants and to cultivate *A. cruentus*, which is a “deep red” form of the plant, as a dye plant (Sauer 1976).

The principal advantage of amaranth is that both the grain and the leaves are sources of high-quality protein. While most grain foods, such as wheat and corn, have 12 to 14 percent protein and lack the essential amino acid lysine, amaranth seeds have 16 to 18 percent protein and are “lysine-rich” (Barrett 1986; Tucker 1986). In areas of the world where animal protein is lacking, the amaranth plant can stave off protein deficiencies. When amaranth flour is mixed with wheat or corn flour in breads or tortillas, the result is a near-perfect protein (eggs which can supply most of the body’s protein requirements). Moreover, amaranth has more dietary fiber than any of the major grains (Tucker 1986). Amaranth seeds also contain calcium and phosphorus, whereas amaranth leaves, which can be eaten like spinach, provide dietary calcium and phosphorus as well as potassium, thiamine, riboflavin, niacin, and vitamins A and C. Amaranth is also richer in iron than spinach (Cole 1979; Tucker 1986).

The amaranth grain (which resembles a miniature flying saucer) is covered with a tough coat that the body cannot digest. Therefore, to obtain nutrition from the seeds it must be processed, and by toasting, boiling, or milling transformed into starch, bran, germ, or oil. Milling the grain yields 28 percent germ-bran and 72 percent white flour. Once processed, the tiny seeds have a “nutty flavor” and are used in breakfast cereals and made into flour for breads (Barrett 1986). Amaranth may also be popped like popcorn or made into candies. The Mexicans mix honey or molasses and popped amaranth into a sweet confection they call “alegría” (happiness) (Marx 1977; Cole 1979). Amaranth may even be brewed into a tea (Barrett 1986).

In addition to its nutritional advantages, amaranth “grows like a weed” in many different environments in the Americas, Africa, and Asia. Although it tends to do best in warm, dry climates with bright sunshine, some species flourish in the wet tropical lowlands, and others do well above 10,000 feet in the Andes. They also tolerate adverse soil conditions, such as high salt, acidity, or alkalinity, in which corn will not survive (Brody 1984; Tucker 1986). Besides growing rapidly under bright sunlight, amaranth has the ability to conserve water by partially closing its leaf pores. It can also tolerate dryness up to a point without wilting (Tucker 1986). Thus, it can be cultivated on marginal soils subject to periodic dry spells, which is an important consideration in semiarid regions.

Two disadvantages of amaranth are that the tiny seeds are awkward to work with and should be harvested by hand. In the United States, machine harvesting is possible after the first severe frost, but yields are lower (Tucker 1986). Yet hand-harvesting is not a major obstacle in countries where agricultural labor is plentiful and its cost is low. Another problem is that the domesticated species easily hybridize via wind pollination with the weedy varieties, yielding low-quality seeds (Marx 1977). Like so many plants, amaranth is also attacked by insects and plant diseases (Tucker 1986). These disadvantages may limit amaranth cultivation to gardeners in the United States and small farmers in the developing world.

Origins

According to J. D. Sauer (1976), wild amaranth seeds were gathered by many Native American peoples. As such, they may have contributed significant protein, as well as essential vitamins, to hunting and gathering populations after the big game animals died out in the Americas. Archaeologists can establish a gradual process of domestication with the appearance of pale, white seeds having improved popping quality and flavor. Notably when seed selection is relaxed, the plants return to producing dark seeds. One of the oldest dates for pale-seeded amaranth is 4000 B.C. from Tehuacan, Puebla, in Mexico, where *A. cruentus* has been found. By 2000 B.C., amaranth was part of the basic Mexican diet (Walsh and Sugiura 1991) (Map II.A.1.1). The Andean species of *A. caudatus* was discovered in 2,000-year-old tombs in northwestern Argentina (Sauer 1976). A more recent date of A.D. 500 marks an additional amaranth species, *A. hypochondriacus*. By the fourteenth century A.D., *A. hypochondriacus* was being cultivated in what is now Arizona.

As Sauer's maps (see also Map II.A.1.2) illustrate, the cores of amaranth cultivation in the pre-Columbian period were in Central Mexico, as well as in the Andes from Peru to northwestern Argentina. Additional pockets were in Ecuador close

to the equator, Guatemala, southern and northwestern Mexico, and southwest North America. By the time the Spanish arrived at Vera Cruz in 1519, amaranth had evolved into a major crop staple employed to satisfy tribute obligations to the Aztec Empire. Moctezuma II received tribute from 17 provinces each year in ivory-white seeds known as *buaubtli* (Sauer 1950), which permitted the Aztecs to fill 18 imperial granaries. According to W. E. Safford (1916) and Sauer (1950), each granary had a capacity of 9,000 to 10,000 bushels. That so many seeds were collected each year is certainly impressive testimony to the widespread cultivation of amaranth in central Mexico before the Spanish conquest.

In addition, the Aztecs raised amaranth on *chinampas* (floating gardens) and utilized the plant in many ways: as a toasted grain, as green vegetables, and as a drink that the Spanish found “delicious.” They also popped it. Since the Aztecs used both the leaves and the seeds, amaranth must have been an important supplement to their diet, especially in times of drought when corn crops died. Why then did the Spanish not adopt such a useful crop? Indeed, not only did the Spanish not adopt amaranth, they actually prohibited it, leaving us to wonder about the extent to which the abolition of such an important source of protein, minerals, and vitamins may have contributed to widespread malnutrition in the sixteenth century.

The Spaniards objected to amaranth because of its ritual uses as a sacred food associated with human sacrifice and “idolatry.” In the early sixteenth century, the Aztecs celebrated a May festival in honor of their patron god, Huitzilopochtli, the god of war, at the great pyramid of Tenochtitlan. The ritual centered on an enormous statue of the god made of amaranth dough and included human sacrifices. Placed on a litter, the statue was carried in procession through the city and then returned to the temple where it was broken up by using other chunks of the same dough. The resulting pieces were subsequently consecrated as the flesh and bones of Huitzilopochtli, then distributed among the people, who ate them with a mixture of reverence and fear. The Spanish missionary and ethnographer Bernardino de Sahagún called the ceremonial paste *zoale* or *tzoalli* and noted that it was also fed to those who were to be sacrificed to Huitzilopochtli (Sauer 1950).

Other deities were also represented by *zoale*, such as the fire god Xiuhtecutli or the goddess Chicomecoatli, but the Tepanecs used it to form bird effigies, and the Tarascans made little figures of animals with the bread of *bledos* (the Spanish term for amaranth). On other occasions, such as the new fire ceremony, this celebration of the new 52-year cycle concluded with everyone present eating the bread of *bledos* and honey (Sauer 1950).



Source: Sauer 1950. Map II.A.1.1. Mexico: Localities and regions where grain amaranth cultivation is indicated.

Despite the Spanish prohibition, however, in the more remote parts of Mexico people continued to cultivate the plant and use it for food and ritual. A half century after the conquest amaranth continued to be an important food throughout much of Mexico (1950). As for ritual in 1629 in Guerrero, the priest Ruiz de Alarcón complained that the Indians were still milling amaranth to make dough for the manufacture of little idols of *zoale* to break up and eat in what appeared to the Spaniards to be a sort of parody of holy communion (Sauer 1950; Early 1992). Even as late as about 1900, a Huichol village of northern Jalisco celebrated one of its major festivals with "cakes" confectioned to represent animals. Made from amaranth seeds mixed with water, these cakes were usually employed ceremonially (Sauer 1950).

Over time, amaranth was also assimilated into Christian rituals. In the late nineteenth century, a visitor to Mexico described rosaries made of little balls of dough that were called *suale* (Sauer 1950). Sauer himself met a woman near Guadalajara in 1947 who was growing grain amaranths, making dough of them, and fashioning them into little cakes and rosaries (Sauer 1950). On a field trip to Mexico and Guatemala, Sauer discovered amaranth being cultivated in many patches, often unknown to outsiders. He found it to be most important in the Federal District, State of Mexico, and Morelos. Other states where it was grown were Guerrero, Tlaxcala, Puebla, Michoacán, and Sonora, Chihuahua, and Sinaloa. Thirty years after Sauer, Daniel Early (1977) visited Tulyehualco in Mexico where he observed techniques of amaranth planting on *chinampas*. In Guatemala, Sauer found that *A. cruentus* was still being planted by the Maya Indians in association with other crops, as they had done in the pre-Columbian period (Sauer 1950, 1976; Morley and Brainerd 1983).

As in highland Guatemala, Indian farmers in the Andes plant amaranth on "the fringes" of their maize fields. Sauer and others have reported amaranth crops in the highlands of Peru and Bolivia and in northwestern Argentina (Sauer 1950). At that time amaranth plants were known by a variety of names: *Achis*, *achita*, *ckoito*, *coyo*, or *coimi* in Peru; and *coimi*, *cuime*, *millmi*, or *quinua millmi* in Bolivia. As Sauer (1950) notes, the term *quinua* was often used for amaranth as well as quinoa.

Amaranth was apparently widely cultivated in the pre-Columbian Andean highlands (Map II.A.1.2). A funerary urn has been found at Pampa Grande in Salta, Argentina, that was full of maize, beans, chenopod seeds, amaranth flowers, and pale seeds identified as *A. caudatus* (Sauer 1950). In 1971, A. T. Hunziker and A. M. Planchuelo reported finding *A. caudatus* seeds in tombs at least 2,000 years old located in northwestern Argentina (Sauer 1976). But in the Inca period, good descriptions of amaranth are lacking, perhaps because it did not play the same ceremonial role as among the Aztecs. The Incas used maize for their

sacred bread rather than amaranth. The first Spanish record that Sauer found (1950) is that of the Jesuit chronicler Bernabé Cobo who reported in 1653 that both red and white *bledos* were commonly consumed by Native Americans. Cobo also recognized that *bledos* were different from quinoa (Sauer 1950).

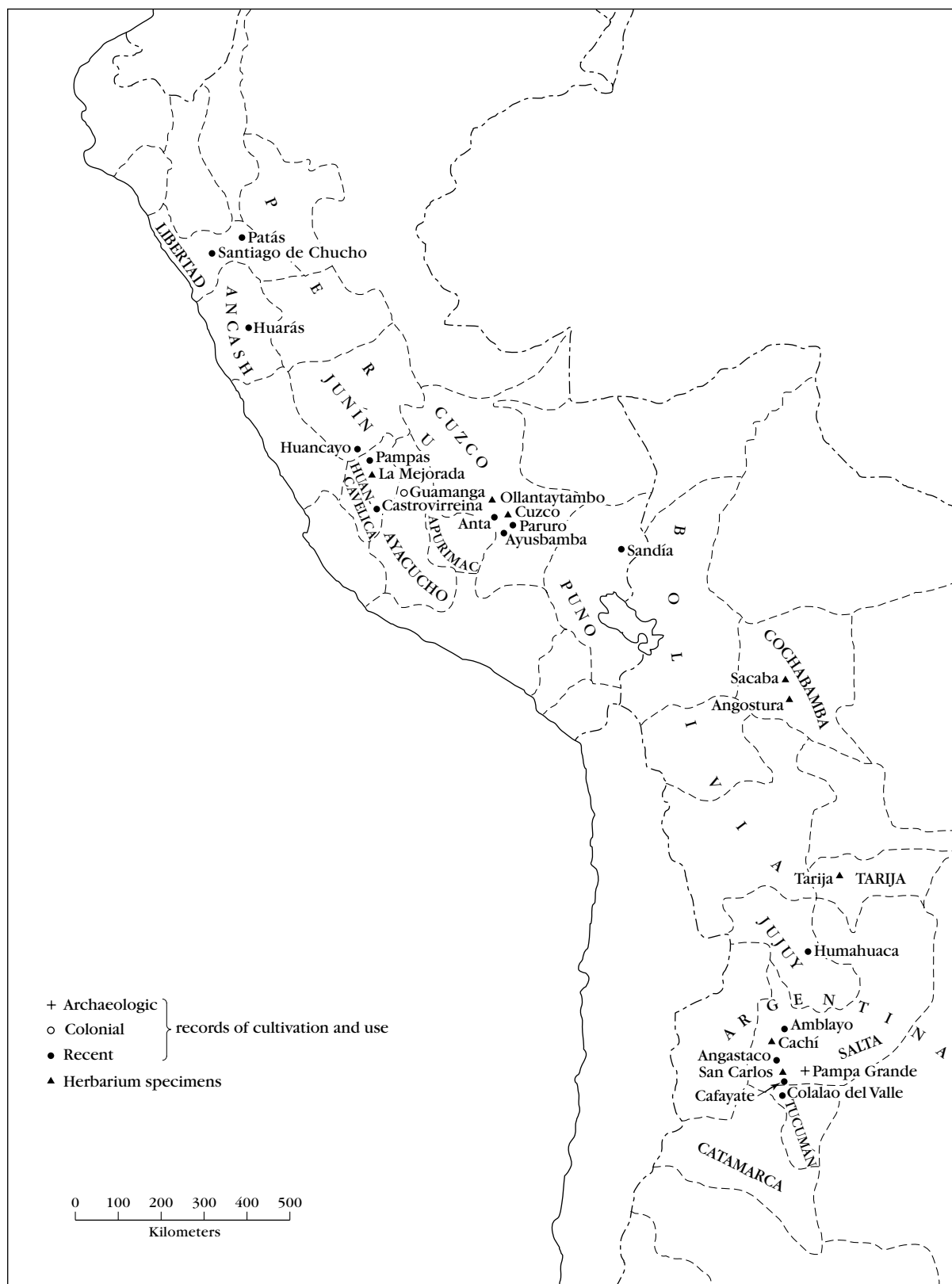
Sometime in the sixteenth century *A. caudatus* was taken from the Andes to Europe. In his *Rariorum Plantarum Historia*, Carl Clusius published the first illustration of the species in Antwerp in 1601 (Sauer 1950). He identified it as *Quinua, sive Blitum majus Peruanum*. Citing Clusius in 1737, Carl von Linné (Linnaeus) named the plant *Amaranthus caudatus* and indicated that it came from South America (Sauer 1950). A pale-seeded variety of *A. hypochondriacus* turned up in a sixteenth-century German collection that was found by P. Hanelt in 1968 (Sauer 1976). According to Sauer (1976), all three domesticated species may have been introduced to the Old World via Europe; henceforth, the American varieties of grain amaranth would be grown as ornamental plants in Europe. Other species of European amaranth now grow wild as weeds.

Asia and North America

The global distribution of amaranth before A.D. 1500 suggests that its cultivation is of "great antiquity," and there is no clear picture of the historical diffusion of American amaranths to Asia, or to Africa, for that matter. Amaranth in Asia probably predates 1492, because it is so widely scattered from Iran to China in remote places, such as Manchuria and eastern Siberia; and it is cultivated by isolated populations in high mountain valleys in the Himalayas. It seems to have been a staple in southern India for centuries (Sauer 1950, 1976), and the Indians argue that it was domesticated in India (Cole 1979).

E. D. Merrill (Cole 1979), however, believed that the Portuguese brought amaranth from Brazil to the Malabar coast of India after 1500, where it was widely cultivated by the nineteenth century. Yet Chinese sources may document its antiquity in Asia. According to Sauer (1950), there apparently is a reference to a grain amaranth in a medical tract of A.D. 950, written for the Prince of Shu in modern Sichuan. It lists six kinds of *bien*, a name that is used for grain amaranths in the same area in modern times.

Most descriptions of cultivated amaranths in Asia are modern, from the nineteenth century on, with a few eighteenth-century references to a European role in diffusing the American plants from Europe to Asia. By the early nineteenth century, amaranth was being cultivated in India, principally in the far south where it was a staple crop in the Nilgiri Hills, and in the north in the foothills of the Himalayas (Sauer 1950, 1976). In the south the people raise it for its seeds, which they convert into flour. Although also grown in the plains regions during the dry winter monsoon, amaranth is especially important in the foothills and mountains of the Himalayas from Afghanistan to Bhutan.



Source: Sauer 1950. Map II.A.1.2. South America: Localities and regions where grain amaranth cultivation is indicated.

Mongoloid nomads on the Tibetan frontier harvest grain at elevations of more than 3,500 meters, while farmers raise *A. caudatus* along with *A. leucocarpus* for grain in Nepal (Sauer 1950). According to Sauer (1950), the Himalayan plants are rich and variable in color with brilliant crimsons and rich yellows (Sauer 1950). Around Tehri in the hills of northern India the grain is popped and made into a dough to form thin cakes. In Nepal the people roast the seeds and eat them in sugar syrup like popcorn balls (Sauer 1950).

Similar popped seeds mixed with a hot sugar syrup are eaten in China, where amaranth seeds are known as *tien-shu-tze*, or millet from heaven. Sauer (1950) received dark-seeded *tien-shu-tze*, grown at elevations of 2,000 to 2,500 meters in northwestern Sichuan and Muping in Sikang. His informant reported that grain amaranths were not grown in Chengtu but in far-off mountain areas. Amaranth was also reportedly grown by the Chinese, who used the seeds to make little cakes (Sauer 1950). In addition, amaranth greens are used for food in China (Zon and Grubben 1976).

In Southeast Asia, amaranth is widely cultivated. In 1747 *A. caudatus* was identified in Indonesia, but the variety widely raised as a vegetable in modern Indonesia and other parts of Asia is *A. tricolor*, as well as species other than *A. caudatus*, such as *A. dubius*. Amaranth is a commercial crop from the Philippines to Taiwan and in Myanmar (Burma), where *A. tricolor* and *A. viridis* are grown (Zon and Grubben 1976).

As A. P. M. van der Zon and G. J. H. Grubben (1976) note, the use of amaranth as a green vegetable is quite extensive in tropical and subtropical regions of Asia and Southeast Asia, where there are many popular names for the plant: *épinard de Chine*, *amarante de Soudan*, African spinach, Indian spinach, *brède de Malabar*, and Ceylon spinach. Its consumption is nearly always in the form of a cooked spinach.

As of the middle 1970s, pale-seeded *A. hypochondriacus* constituted the bulk of the Asiatic crop; dark-seeded *A. hypochondriacus* and pale-seeded *A. caudatus* were minor components, although their leaves were widely used as a vegetable (Sauer 1976; Zon and Grubben 1976). The third American species, *A. cruentus*, has generally been planted as an ornamental dye plant or herb in Asia. Amaranth cultivation has been spreading in India, and Sauer (1976) believed at that time that India was the one place where amaranth was likely to experience expanded cultivation, perhaps stimulated by plant breeding. This was, of course, before scientists in the United States began extensive research on grain amaranths in the late 1970s and 1980s and American farmers initiated commercial production of amaranth in 1983 (Robinson n.d.).

The Rodale Research Center in Kutztown, Pennsylvania, is one of the leading amaranth research centers in the United States. In part because of amaranth's seed distribution and partly because of promotional efforts, American small farmers are increasingly cultivating the grain. Breads, cereals, cookies, and "Gra-

ham" crackers made with amaranth are now available in health-food and grocery stores in the United States.

Africa

How and when amaranth reached Africa is also uncertain. It is widely found as an ornamental plant or weed throughout the continent: from Senegal to Nigeria in West Africa, from Equatorial Africa to Zaire; and to a lesser extent in East Africa. Species have even been identified in Morocco, Ethiopia, and Sudan (Sauer 1950). *A. cruentus* is widespread in Africa, and *A. caudatus* is used as an ornamental and grain amaranth.

According to Zon and Grubben (1976), a variety of *A. cruentus* was introduced "recently" to southern Dahomey (now Benin) from Suriname and has proved to be resistant to drought. Other varieties of *A. cruentus* were widely marketed as a vegetable crop in Dahomey or raised in family gardens. A third American amaranth, *A. hypochondriacus*, was introduced into East Africa in the 1940s as a grain for the Indian population there (Sauer 1976).

That the history of amaranth in Africa remains comparatively unknown may be because of its widespread use in private vegetable gardens. The food preference of many Africans and African-Americans for greens may also mean that amaranth leaves are more important in the diet of Africans than seeds. The problem for the historical record, however, is that the cultivation and consumption of greens usually escapes documentation.

Conclusion

At this stage of research, the historical evolution and diffusion of amaranth remains a research problem for the future. How did amaranth come to be cultivated around the world? Under what conditions and when? As Sauer (1950) notes, amaranth species, cultivation methods, and consumption patterns are remarkably similar in Old and New Worlds. In both areas amaranth tends to be cultivated in the highlands, although the weed species grow well at other altitudes.

Farmers usually cultivate amaranth in conjunction with maize and other crops and consume it themselves in the form of balls of popped seeds, meal, and little cakes, and they use the seeds to make a beverage. It was and is a food crop principally of interest to small farmers and gardeners but one that promises to resolve some problems of world hunger in the twenty-first century. The great advantage of the grain amaranths is that they can be grown on marginal soils where reliable water supplies are problematic and can nourish populations that lack access to animal protein.

Once widely cultivated as a staple crop of the Aztec empire, amaranth has already proven its ability to sustain millions of people in a region lacking in significant sources of animal protein long before the arrival of the

Europeans. If population growth forces small farmers to move to more marginal lands in the next century, amaranth may make an important difference in nutrition levels for people who lack access to good corn or rice lands and cannot afford meat. In short, amaranth may well become an important supplementary food crop in many developing countries in the future.

Mary Karasch

Bibliography

- Amaranth. 1989. *Academic American Encyclopedia*, Vol. 1, 321. Danbury, Conn.
- Amaranthus. 1991. In *The Encyclopedia Americana International Edition*, Vol. 1, 653. Danbury, Conn.
- Barrett, Mariclaire. 1986. The new old grains. *Vegetarian Times* 101: 29-31, 51.
- Berberich, Steven. 1980. History of amaranth. *Agricultural Research* 29: 14-15.
- Brody, Jane E. 1984. Ancient, forgotten plant now "grain of the future." *The New York Times*, Oct. 16.
- Cole, John N. 1979. *Amaranth from the past for the future*. Emmaus, Pa.
- Cramer, Craig. 1987. The world is discovering amaranth. *The New Farm* 9: 32-5.
- Early, Daniel. 1977. Amaranth secrets of the Aztecs. *Organic Gardening and Farming* 24: 69-73.
1992. The renaissance of amaranth. In *Chilies to chocolate: Food the Americas gave the world*, ed. Nelson Foster and Linda S. Cordell, 15-33. Tucson, Ariz.
- Gates, Jane Potter. 1990. *Amaranth for food or feed January 1979-December 1989. Quick bibliography series: QB 90-29. Updates QB 88-07. 210 Citations from Agricola*. Beltsville, Md.
- Hunziker, Armandu T. 1943. Las especies alimenticias de Amaranthus . . . cultivados por los indios de América. *Revista Argentina de Agronomía* 10: 297-354.
- Marx, Jean. 1977. Amaranth: A comeback for the food of the Aztecs? *Science* 198: 40.
- Morley, Sylvanus G., and George W. Brainerd. 1983. *The ancient Maya*. Fourth edition, rev. Robert J. Sharer. Stanford, Calif.
- Robinson, Robert G. n.d. Amaranth, quinoa, ragi, tef, and niger: Tiny seeds of ancient history and modern interest. *Agricultural Experiment Station Bulletin*. St. Paul, Minn.
- Safford, W. E. 1916. An economic amaranthus of ancient America. *Science* 44: 870.
- Sauer, J. D. 1950. The grain amaranths: A survey of their history and classification. *Annals of the Missouri Botanical Garden* 37: 561-632.
1976. Grain amaranths. In *Evolution of crop plants*, ed. N. W. Simmonds, 4-7. London and New York.
- Tucker, Jonathan B. 1986. Amaranth: The once and future crop. *BioScience* 36: 9-13.
- Walsh, Jane MacLaren, and Yoko Sugiura. 1991. The demise of the fifth sun. In *Seeds of change*, ed. Herman J. Viola and Carolyn Margolis, 16-44. Washington, D.C., and London.
- Wister, John C. 1985. Amaranth. In *Collier's Encyclopedia*, Vol. 1: 622. London and New York.
- Zon, A. P. M. van der, and G. J. H. Grubben. 1976. *Les légumes-feuilles spontanés et cultivés du Sud-Dabomey*. Amsterdam.

II.A.2. 🌾 Barley

That people do not live "by bread alone" is emphatically demonstrated by the domestication of a range of foodstuffs and the cultural diversity of food combinations and preparations. But even though many foods have been brought under human control, it was the domestication of cereals that marked the earliest transition to a food-producing way of life. Barley, one of the cereals to be domesticated, offered a versatile, hardy crop with an (eventual) tolerance for a wide range of climatic and ecological conditions. Once domesticated, barley also offered humans a wide range of valuable products and uses.

The origins of wheat and barley agriculture are to be found some 10,000 years ago in the ancient Near East. Cereal domestication was probably encouraged by significant climatic and environmental changes that occurred at the end of the glaciated Pleistocene period, and intensive harvesting and manipulation of wild cereals resulted in those morphological changes that today identify domesticated plants. Anthropologists and biologists continue to discuss the processes and causes of domestication, as we have done in this book's chapter on wheat, and most of the arguments and issues covered there are not reviewed here. All experts agree, however, on the importance of interdisciplinary research and multiple lines of evidence in reconstructing the story of cereal domestication.

Readers of this chapter may note some close similarities to the evidence for wheat domestication and an overlap with several important archaeological sites. Nonetheless, barley has a different story to tell. Barley grains and plant fragments are regular components of almost all sites with any plant remains in the Near East, regardless of period or food-producing strategy. Wild barley thrives widely in the Near East today - on slopes, in lightly grazed and fired pastures, in scrub-oak clearings, in fields and field margins, and along roadsides. These circumstances suggest a different set of research questions about barley domestication, such as: What was barley used for? Was its domestication a unique event? And how long did barley domestication take?

In addition, there are subthemes to be considered. Some researchers, for example, have suggested that barley was not domesticated for the same reasons that led to the domestication of other cereals - such as the dwindling of other resources, seasonal shortages, a desire for a sedentary food base, or the need



Barley

for a surplus for exchange. Instead, barley may have been cultivated for the brewing of ale or beer.

In another view, the very slight differences between the wild and domesticated forms, and the ease with which wild barley can be domesticated, make it difficult to believe that barley domestication did not occur more than once. Geneticists and agricultural historians have generally believed that groups of crops were domesticated in relatively small regions and spread by human migration and trade. If barley was domesticated independently in several different communities, this would indicate that the transition to farming in those areas required little innovation and took place under recurring conditions.

Finally, because of their presence in many excavated sites, ancient barleys provide some of the best archaeological evidence bearing on the general problem of the pace of plant domestication. Whether the process took place over the course of a single human lifetime or evolved over many decades or centuries remains an important issue that may never be resolved with archaeological evidence alone (Hillman and Davies 1990). But the pace of domestication lies at the heart of the debate over the Neolithic – was it revolution or evolution (Childe 1951; Rindos 1984)? Did plant domestication radically transform people's lifestyles, or was it a gradual by-product of long-term behaviors with radical consequences noted only in retrospect? Barley is a crop that may hold answers to such questions.

Archaeological Evidence for the Domestication of Barley

Archaeological evidence points to the domestication of barley in concert with the emergence of Neolithic villages in the Levantine arc of the Fertile Crescent. Pre-Neolithic peoples, notably Natufian foragers (whose cultural remains include relatively large numbers of grinding stones, sickle blades, and storage pits), increasingly depended on plant foods and, perhaps, plant cultivation. Unfortunately, their tools point only to general plant use, and archaeologists continue to discuss which plants were actually processed (e.g., McCorrison 1994; Mason 1995). There is no evidence to suggest that the Natufians domesticated barley or any other plant.

Charred plant remains, the best evidence for domestication of specific plants, have rarely been recovered from pre-Neolithic sites. Preservation is generally poor. In the case of barley, only a few sites prior to the Neolithic contain recognizable fragments, and all examples indicate wild types. A few barley grains from Wadi Kubbania, an 18,000-year-old forager site in southern Egypt, were at first thought to be early examples of domesticated barley (Wendorf et al. 1979), but subsequent laboratory tests showed these to be modern grains that had intruded into ancient occupation layers (Stemler and Falk 1980; Wendorf et al. 1984). Other plant remains from Wadi Kubbania

included relatively high numbers of wild *Cyperus* tubers and wild fruits and seeds (Hillman 1989).

One of the most extraordinary pre-farming archaeological sites to be discovered in recent years is Ohalo II, on the shore of the Sea of Galilee, which yielded quantities of charred plant remains, including hundreds of wild barley grains (Kislev, Nadel, and Carmi 1992). About 19,000 years ago, foragers camped there, and the remains of their hearths and refuse pits came to light during a phase of very pronounced shoreline recession several years ago. Excavators believe that the charred plants found in the site were the remains of foods collected by Ohalo II's Epi-Paleolithic foragers. If so, these foragers exploited wild barley (*Hordeum spontaneum* Koch.), which was ancestral to domesticated barley.

Despite the generally poor preservation of plant remains, there are two Natufian sites with evidence suggesting that foraging peoples there collected some wild barley just prior to the beginnings of agriculture. Wadi Hammeh, a 12,000-year-old Early Natufian hamlet overlooking the Jordan Valley (Edwards 1988), contained charred seeds of wild barley (*Hordeum spontaneum*), wild grasses, small legumes, crucifers, and a range of other plants, as yet unidentified (Colledge, in Edwards et al. 1988). The seeds were scattered among deposits in several round houses, somewhat like the scatter of plant remains at another Natufian site, Hayonim Cave, where Early and Late Natufian dwellers had constructed round houses, possibly seasonal dwellings, within the cave (Hopf and Bar Yosef 1987). To be certain that disturbances had not carried later seeds down into Natufian layers (a problem at Wadi Kubbania and at another Natufian site, Nahel Oren), excavators had the charred seeds individually dated. Wild lupines found with wild almonds, wild peas, and wild barley (*Hordeum spontaneum*) suggest that the Natufian inhabitants collected plants that could be stored for later consumption.

Although there is little doubt that some foraging groups collected wild barley, evidence for the beginnings of barley domestication is far more controversial. Until recently, archaeologists were convinced that farmers (as opposed to foragers) had occupied any site containing even a few barley grains or rachis fragments with the morphological characteristics of domestic cereals. Thus, the identification of tough-rachis barley in the Pre-Pottery Neolithic A (PPNA) levels at Jericho (Hopf 1983: 609) implied that the earliest Neolithic occupants domesticated barley in addition to wheats and legumes.¹ A tough rachis inhibits seed dispersal, and wild barleys have brittle rachises that shatter when the seeds mature. Each segment of the rachis supports a cluster of three florets (flowers), from which only one grain develops. If a tough rachis fails to shatter, the seeds remain on the intact stalk, vulnerable to predators and unable to root and grow. Yet a tough-rachis crop is more easily and efficiently harvested by humans. Through human

manipulation, the tough-rachis trait, which is maladaptive and scarce in the wild (Hillman and Davies 1990: 166–7), dominates and characterizes domesticated barley. At Jericho, the tough-rachis finds suggested to archaeologists that barley domestication had either preceded or accompanied the evident Neolithic practices of plant cultivation using floodwater manipulation in an oasis habitat.

However, at Netiv Hagdud, a contemporary (PPNA) site to the north of Jericho, Neolithic settlers seem to have practiced barley cultivation (Bar-Yosef et al. 1991), and the recovery there of rich archaeobotanical remains has forced archaeobotanists to rethink the significance of a few barley remains of the domesticated type in Early Neolithic sites throughout the Near East. Built on an alluvial fan in a setting not unlike that of Jericho, Netiv Hagdud proved to contain the foundations of almost a dozen large oval and small round structures. Some of these were probably houses with rock platform hearths and grinding equipment (Bar-Yosef et al. 1991: 408–11). The charred plant remains from the site included thousands of barley grains and rachis fragments, and the opportunity to examine these as an assemblage led to a surprising discovery.

Although a number of fragments clearly displayed a domesticated-type tough rachis (Kislev and Bar-Yosef 1986), archaeobotanists realized that as an assemblage, the Netiv Hagdud barley most closely resembles modern *wild* barley. Even among a stand of wild barley, approximately 12 percent of the spikes have a tough rachis (Zohary and Hopf 1993: 62) because this characteristic regularly appears as the result of mutation and self-fertilization to generate a pair of recessive alleles in a gene (Hillman and Davies 1990: 168). At Netiv Hagdud, the thousands of barley remains (including low numbers of tough-rachis examples) were collected, possibly from cultivated stands, by Early Neolithic people who appear from available evidence to have had no domesticated crops, although some scholars suggest that harvesting timing and techniques might still make it difficult to distinguish between wild and semidomesticated barley (Kislev 1992; Zohary 1992; Bar-Yosef and Meadow 1995). That evidence also implies that occasional domesticated-type barley remains at other Neolithic sites, including Tell Aswad (van Zeist and Bakker-Heeres 1982), may actually belong to an assemblage of purely wild barley.

Other Early Neolithic sites with remains of barley include Gilgal I, a PPNA site with cultural remains similar to those at Netiv Hagdud and Jericho. It is still unclear whether the “large amounts of oat and barley seeds” recovered in the mid-1980s from a silo in one of the Neolithic houses were domesticated or wild types (Noy 1989: 13). Tell Aswad, formerly nearer to Mediterranean oak forests and set beside a marshy lakeshore in the Damascus Basin, has also yielded a mix of rachis fragments, predominantly wild-type but

with some domesticated-type (van Zeist and Bakker-Heeres 1982: 201–4). This same mix of wild and domesticated types was found in early levels at other early sites, including Ganj Dareh on the eastern margins of the Fertile Crescent (van Zeist et al. 1984: 219).

Over time, the percentages of wild-type and domesticated-type barley remains were inverted at sites in the Damascus Basin (van Zeist et al. 1984: 204). Tell Aswad was the earliest occupied of these sites in a 2,000-year sequence of nearly continuous residence there (Contenson 1985), and when it was abandoned, farmers had already settled nearby, at Tell Ghoraifé. Although the barley remains at Ghoraifé included many domesticated-type specimens, it was from evidence of a later occupation at Tell Aswad (8,500 to 8,900 years ago) and the nearby site of Ramad (occupied from 8,200 years ago) that archaeobotanists could unequivocally distinguish between fully domesticated barley and low numbers of wild barley in the same assemblages. The archaeobotanists suggested that the apparent mix of wild and domesticated types in earlier deposits may indicate an intermediate stage in the domestication process (van Zeist and Bakker-Heeres 1982: 184–5, 201–4).

There has never been much doubt that remains of barley from northern Levantine sites indicate plant collection or incipient stages of cultivation without domestication 10,000 years ago. Wild barley (*Hordeum spontaneum* Koch.) is well represented among the plant remains from Mureybit, an Early Neolithic village on the Euphrates River in the Syrian steppe (van Zeist and Bakker-Heeres 1984a: 171). Contemporary levels at Qeremez Dere, also in the steppe, contain abundant remains of fragmented wild grass grains, some of which “seem most probably to be wild barley” (Nesbitt, in Watkins, Baird, and Betts 1989: 21). Plant remains suggest that cereal cultivation was of little importance as one moved further north and east during the first centuries of Neolithic occupation. Cereals were present only as traces in the northern Iraq site of Nemrik 9 along the Tigris River (Nesbitt, in Kozłowski 1989: 30), at M’lefaat (Nesbitt and Watkins 1995: 11), and not at all at Hallan Çemi (Rosenberg et al. 1995) nor at earliest Çayönü (van Zeist 1972).

Available evidence now seems to suggest that domesticated barley appeared in the second phase of the early Neolithic – Pre-Pottery Neolithic B (PPNB) – several hundred years after wheat farming was established.² By the PPNB (beginning around 9,200 years ago), several different forms of domesticated barley, two-row and six-row (see the section “Taxonomy”), appeared among plant remains from Neolithic sites. Jericho and Ramad had both forms from about 9,000 years ago (van Zeist and Bakker-Heeres 1982: 183; Hopf 1983: 609). Barley does not seem to have been among the earliest crops in the Taurus Mountains – at PPNB sites such as Çayönü (van Zeist 1972; Stewart

1976) and Çafar Höyük (Moulins 1993). At Neolithic Damishliyya, in northern Syria (from 8,000 years ago), and Ras Shamra on the Mediterranean coast (from 8,500 years ago), domesticated two-row barley was present (van Zeist and Bakker-Heeres 1984b: 151, 159; Akkermans 1989: 128, 1991: 124). In Anatolia, domesticated barley also characterized the first appearance of farming, for example, at Çatal Hüyük (Helbaek 1964b). This is also the case in the eastern Fertile Crescent, where domesticated plants, barley among them, first showed up around 9,000 years ago – at Jarmo (Helbaek 1960: 108–9) and Ali Kosh (Helbaek 1969).

An excellent review of archaeobotanical remains (Zohary and Hopf 1993: 63–4) tracks the subsequent spread of two-row and six-row barley around the Mediterranean coast, across the Balkans, and into temperate Europe. Barley was one of the fundamental components of the Neolithic economic package introduced (and modified) with the spread of Near Eastern farmers across Anatolia and into various environmental zones of Europe (Bogucki 1996), as well as into northern Egypt (Wetterstrom 1993), Central Asia (Charles and Hillman 1992; Harris et al. 1993), and South Asia. By 8,000 years ago, barley agriculture had reached the foothills of the Indus Valley, where it supported farmers at Mehrgarh, one of the earliest settlements in South Asia (Jarrige and Meadow 1980; Costantini 1984).

During these first several thousand years, domesticated barley also spread into the steppes and desert margins of the Near East, expanding farming practices into an ecological niche where risk was high and barley offered what was probably the best chance for crop survival. At Abu Hureyra, in the Syrian steppe near Mureybit, six-row barley appeared earlier than its two-row barley ancestor (Hillman 1975), suggesting that fully domesticated barley crops were introduced to that site (Hillman and Davies 1990: 206), although recent paleoecological models suggest that wild barley grew nearby (Hillman 1996).

By the end of the PPNB, sites with barley were fairly common in very marginal farming zones – including, for example, El Koum 2 (Moulins, in Stordeur 1989: 108), Bouqras (van Zeist and Waterbolk-van Rooijen 1985), and Beidha (Helbaek 1966). At Nahal Hemar cave, in the dry Judean Hills, a single kernel of domesticated barley recovered was presumably imported to the site some 9,000 years ago (Kislev 1988: 77, 80). The spread of barley not only indicates the success of a farming way of life but also offers important archaeological indicators to complement botanical and ecological evidence of the domestication and significance of this vital crop plant.

Botanical Evidence for the Domestication of Barley

Barley grains and barley rachis internodes recovered from archaeological sites show the telling morphological characteristics of wild and domesticated forms,

but botanical studies of modern barleys have also provided critical evidence for an interdisciplinary reconstruction of barley domestication. Archaeologists can examine ancient morphology but not ancient plant behavior. It was changes in both characteristics that established barley as a domesticated plant. Not only did the rachis become tough, but the ripening period of barley narrowed, and a single-season seed dormancy became established. Botanical studies have revealed relationships between species and varieties of barley and the precise nature of the changes that must have occurred under domestication.

Taxonomy

As with wheats, taxonomic classification of barleys has changed with expanding scientific knowledge of genetic relationships. Genetic evidence now indicates much closer relationships among what morphologically once appeared to be distinctive species (Briggs 1978: 77). Yet it is the morphological criteria, easily seen, that offer farmers and archaeologists a ready means by which to classify barleys (but see Hillman and Davies 1990, and Hillman et al. 1993, for alternative experimental approaches). Because archaeologists have had to rely largely on morphological criteria to detect the beginnings of domestication, the old species names remain convenient terms for distinguishing between what are now considered barley varieties (new species names in parentheses follow).

Barleys belong in the grass tribe Triticeae, to which wheats and ryes (barley's closest crop relatives) also belong. There are 31 barley species (almost all wild), and nearly three-fourths of them are perennial grasses (Bothmer and Jacobsen 1985; Nilan and Ullrich 1993). Despite the diversity of wild barleys that can be identified today, most botanists and geneticists concur that all domesticated types most probably have a single wild ancestor, *Hordeum spontaneum* Koch. (*H. vulgare* subsp. *spontaneum*) (Harlan and Zohary 1966; Zohary 1969). This plant crosses easily with all domesticated barleys. The major morphological difference between wild and domesticated barley lies in the development of a tough rachis in the domesticate.

Once farmers had acquired a domesticated two-row form of barley, they selectively favored the propagation of a further morphological variant, the six-row form. Barleys have three flowers (florets) on each rachis segment (node). In the wild and domesticated two-row forms, however, only the central floret develops a grain. Thus, one grain develops on each side of a rachis, giving the spike the appearance of two grains per row when viewed from the side. In the six-row form, *Hordeum hexastichum* L. (*H. vulgare* subsp. *vulgare*), the infertility of the lateral florets is overcome: Nodes now bear three grains each, so the spike has three grains on each side. This gives an appearance of six grains per row in side view. A general evolutionary trend in the grass family has been the reduction of reproductive parts; so, for a long time, it was

difficult for botanists to accept that one of the consequences of domestication and manipulation of barley has been to restore fertility in lateral spikelets, thereby increasing the grain production of each plant (Harlan 1968: 10).

A final important morphological change was the appearance of naked-grain barleys. In wild cereals, the modified seed leaves (glumes, lemmas, and paleas) typically tightly enclose the grain and form a protective husk. From a human perspective, one of the most attractive changes in domesticated cereals is the development of grains from which the glumes, lemmas, and paleas easily fall away. Because humans cannot digest the cellulose in the husks, such development considerably reduces processing effort. Naked-grain barleys (*Hordeum vulgare* subsp. *distichum* var. *nudum* and *H. vulgare* subsp. *vulgare* var. *nudum*) appeared shortly after the emergence of six-row forms (Zohary and Hopf 1993: 63). Taxonomists have always recognized these as varieties rather than as distinct species of barley.

Genetics

Genetic relationships remain very close among all barleys, and modern taxonomic schemes collapse all cultivated barleys and the wild *Hordeum spontaneum* ancestor into a single species, *Hordeum vulgare* (Harlan and Zohary 1966: 1075; Briggs 1978: 78; Nilan and Ullrich 1993: 3). *H. vulgare* is a diploid with two sets of seven chromosomes ($2n = 14$) and has proved an excellent subject for genetic and cytogenetic analysis (Nilan and Ullrich 1993: 8). Because the plant is self-fertile (that is, male pollen fertilizes its own or adjacent flowers on the same plant), mutations have a good chance of being copied and expressed in the genes of subsequent generations. This attribute was undoubtedly an important feature of barley domestication, for a very few mutations have caused major morphological changes that were easily favored by humans, both consciously and unconsciously (Harlan 1976; Hillman and Davies 1990).

A brittle rachis, for example, is controlled by a pair of tightly linked genes. A mutant recessive allele in either gene (Bt and Bt_1) will produce a tough rachis in homozygous offspring (Harlan 1976: 94; Briggs 1978: 85). This condition will occur rarely in the wild but may be quickly selected for and fixed in a population under cultivation (Hillman and Davies 1990: 166–8). Experimental trials and computer simulations suggest that under specific selective conditions, the homozygous recessive genotype may become predominant in as few as 20 years (Hillman and Davies 1990: 189). Consequently, barley domestication depends on one mutation!

Furthermore, a single recessive mutation also is responsible for fertility in lateral florets and the conversion from two-row to six-row forms (Harlan 1976: 94). Another gene, also affected by a recessive mutant

allele, controls the adherence of lemma and palea to the grain. Jack Harlan (1976: 95–6) has postulated a single domestication of wild barley followed by other recessive mutants for six-row and naked forms.

Objections to this parsimonious reconstruction revolve around the many brittle or semibrittle rachis variants of barley, some of which include six-row forms. Barley is rich in natural variants in the wild (Nevo et al. 1979; Nilan and Ullrich 1993: 9), and geneticists have long tried to incorporate six-row brittle forms (e.g., *Hordeum agriocritibon* Åberg) into an evolutionary taxonomy of the barleys. Most now agree that the minor genetic differences, wide genetic diversity, and ease of hybridization and introgression with domesticated forms account for the great number of varieties encountered in the “wild” (Zohary 1964; Nilan and Ullrich 1993: 3).

Ecological Evidence for Barley Domestication

Geographic Distribution

In the tradition of Nikolay Ivanovich Vavilov, geneticists and botanists have documented the distributions of wild and domesticated varieties of barley. The places where wild races ancestral to domesticated crops grow today may indicate the range within which a crop arose, because the earliest cultivators must have encountered the plant in its natural habitat. Patterns of genetic diversity in different areas may also offer clues about the history of a crop plant. Such patterns may suggest, for example, a long and intensive history of manipulation or an isolated strain. Harlan and Daniel Zohary (1966: 1075–7) have summarized the modern distribution of wild barley, *Hordeum spontaneum*, noting many populations outside the core area of the Fertile Crescent where the earliest agricultural villages lay.

Harlan and Zohary distinguish between truly wild barley (*H. spontaneum*) and weedy races – wild-type barleys derived from domesticated barley crops in areas to which barley farming spread after domestication. Modern distribution of truly wild progenitors is closely associated with the geography of semiarid Mediterranean climates and with an ecological relationship with deciduous oak open woodland that covers the lower slopes of the mountain arc of the Fertile Crescent. This landscape was made famous by Robert J. Braidwood’s expedition to the “Hilly Flanks,” where he and later archaeologists sought to uncover the first farming villages (Braidwood and Howe 1960: 3). A “small, slender, very grassy type” grows wild in steppic environments with somewhat greater temperature extremes and less annual rainfall. Another distinct truly wild race in the southern Levant has extremely large seeds (Harlan and Zohary 1966: 1078).

In areas outside the semiarid Mediterranean woodlands, wild-type barleys survive in a more continental climate (hotter summers, colder winters, year-round

rainfall). Because all the collections in such areas have proved to be weedy races (including brittle-rachis, six-row *Hordeum agriocrithon* Åberg from Tibet), their range provides better information on the spread of barley farming than on the original domestication of the plant.

Ecological Factors

Ecologically, wild barley shares some of the preferences of wheat, but wild barley has not only a much more extensive geographical range but also a wider ecological tolerance. Barleys thrive on nitrogen-poor soils, and their initial cultivation must have excluded dump-heap areas enriched by human and animal fertilizers (Hillman and Davies 1990: 159). But the wild barley progenitors do thrive in a variety of disturbed habitats (Zohary 1964). In prime locales, wild barley flourishes on scree slopes of rolling park-woodlands. It likes disturbed ground – in fields and along roadsides – and is a moderately aggressive fire follower (Naveh 1974, 1984).

Ecological and botanical attributes of wild barley have convinced some that it was the first domesticate (Bar-Yosef and Kislev 1989: 640). Archaeological evidence, however, indicates that despite possible cultivation in the PPNA (Hillman and Davies 1990: 200; Bar-Yosef et al. 1991), barley was not domesticated as early as wheat and some legumes. From the perspective of cultivators, wild wheats have several advantages over wild barley, including greater yield for harvesting time (Ladizinsky 1975) and easy detachment of lemma and palea. Modern wild barleys demonstrate a number of features potentially attractive to foraging peoples, including large grain size, ease of mutant fixation, local abundance, and wide soil and climate tolerance (Bar-Yosef and Kislev 1989: 640). Nevertheless, the archaeological record holds earlier evidence of domesticated wheat than of domesticated barley.

Was wheat domestication fast and that of barley slow? Or were these cereals domesticated at the same rate but at different times? Experimental studies offer significant insights into the archaeological record of cereal domestication, including probable causes of ambiguity where wild forms may or may not have been cultivated (Hillman and Davies 1990; Anderson-Gerfaut, Deraprahamian, and Willcox 1991). Although barley domestication can happen very quickly, the rate of domestication would have varied according to planting and harvesting conditions. It would be nearly impossible to discriminate between collection and cultivation (reseeding) of wild barley in the archaeological record; therefore, it is difficult to know whether barley was cultivated for a long time before becoming a recognizable domesticate. The excellent work of Gordon Hillman and Stuart Davies (1990) with mutation rates, cultivation variables, and harvest strategies suggests little inherent difference between wheat and barley plants for their domesticability. Per-

haps the very different archaeological record, with a PPNA emergence of domesticated wheat and a later PPNB emergence of domesticated barley, implies that we should look beyond the genetics and ecology of the plants for other variables in the early history of agriculture.

Uses of Barley and Barley Domestication

Today, barley is primarily important for animal feed, secondarily for brewing beer, and only marginally important as a human food. Although researchers typically assume that cereals were first domesticated as foodstuffs, we do not know what prominence, if any, barley had in early agriculturalists' diets. In its earliest, most primitive, hulled form, barley required extra processing to remove lemma and palea. Once naked barleys appeared, one suspects that they were preferred as human food, but one cannot conclude that hulled barleys were domesticated for animal feed. Although the domestication of barley coincided with the appearance of domesticated animals in the Levant, the Natufian and PPNA evidence clearly indicates harvest and, perhaps, cultivation of wild barley before animal domestication. Some quality of barley attracted cultivators before they needed animal feed.

Solomon Katz and Mary Voigt (1986) have hypothesized that barley domestication was a consequence of early beer brewing. They suspect that epi-Paleolithic peoples intensively cultivated wild barley because they had come to understand its use in fermentation and the production of alcohol, and it was this use that prompted the advent of Neolithic farming. Their theory implies that epi-Paleolithic peoples were sedentary, as Natufians apparently were (Tchernov 1991). Beer brewers must also have possessed pottery or other suitable containers, yet the invention of pottery took place long after cereal domestication in the Near East. And if cereal (barley) domestication was brought about by demand for beer, then domestication probably was impelled by social relationships cemented by alcohol rather than by subsistence values of cereal grains. The social context of drinking has been explored recently by a number of anthropologists (e.g., Moore 1989; Dietler 1990) who have emphasized the important roles that alcohol and other costly perishables play in social relationships, especially in matters of reciprocity and obligation.

Perhaps one of the most significant insights into the theory that a fermented beverage (rather than a nutritive grain) impelled domestication lies in the methods by which early beer was made. Recipes on Mesopotamian clay tablets and iconographic documentation clearly indicate that early beer was made from bread (Katz and Maytag 1991) rather than from malt (sprouted grain), as in modern practice. Both bread making and malting produce a fermentation material in which the cereal-grain endosperm has already been partially broken down mechanically and

chemically (Hough 1985: 4–5). Archaeologists have detected residues consistent with beer making on ceramics from as early as the late fourth millennium B.C. (Michel, McGovern, and Badler 1992). If this Sumerian beer-making tradition developed from early antiquity, the most parsimonious theory is that beer-making developed from fermented bread.

But even if barley was first cultivated for its food value (perhaps as grits or gruel), it clearly offered an important array of products to farmers. In addition to grain feed, beer, and bread, barley also yielded straw for fodder, thatch, basketry, mudbrick, and pottery temper. Many of these products must have been essential as barley farming spread to arid regions all but devoid of trees, wood, and lush, wild vegetation.

The Spread of Barley Farming

Perhaps the most significant expansion of domesticated barley was into the truly arid steppes and deserts of the Near East, where irrigation was critical to its survival. Hans Helbaek (1964a: 47) has argued that barley was irrigated by the occupants of Tell es Sawwan in the early fifth millennium B.C. Six-row barley was evident among the site's archaeological plant remains, yet with the available rainfall, plants producing even a third as much seed (such as the two-row forms) would have been hard-pressed to survive; the six-row form could have thrived only under irrigation. Six-row barley was one of the principal crops in ancient southern Mesopotamia, and some have even suggested that barley production lay at the heart of the rise of the earliest truly complex societies in a river-watered desert.

Barley farming has also expanded into temperate regions of China, and into the tropics, where dry and cool highland regions (such as in Ethiopia, Yemen, and Peru) offer appropriate locales for cultivation. Christopher Columbus's voyages first brought the crop to the New World (Wiebe 1968), where it spread most successfully in North America (Harlan 1976: 96).

Conclusion

Domesticated barley was an important crop in human prehistory and provided many products to a wide range of settled peoples. Barley has long been considered one of the initial domesticates in the Southwest Asian Neolithic package of cereals and legumes, and in a broad chronological sweep, this conclusion remains true. But recent archaeological and botanical research indicates that domesticated barley did not appear among the very first domesticated plants. With a growing corpus of plant remains, more attention can be paid to regional variation in the Southwest Asian Neolithic, and archaeologists can now develop a more complex understanding of early plant domestication. Just as there seems to be no barley in PPNA agriculture, barley also seems to be

absent from early PPNB farming communities in the Taurus Mountains. Archaeologists have long recognized other late domesticates, such as grapes, olives, and other fruits. The progress of barley domestication, along with its origins, offer potentially interesting insights into the development of domestic lifestyles and the expansion and adoption of farming in new ecological zones.

One explanation for different timing of wheat and barley domestication might be found in the possibility that differences in cultivation practices led to differences in cereal domestication. Modern studies suggest that domestication rates should be similar for wheat and barley, but they also demonstrate that different practices – tending cereals on the same or on new plots each year, for example – can affect domestication rates. An archaeological record implying a long period of cultivation for barley prompts us to wonder if cultivators treated wheats and barleys differently. Did they value one cereal more than another? Can we use such evidence from different regions and different crops to understand the significance of different plants in the diets and lives of the earliest cultivators and farmers?

Increasing botanical and ecological knowledge of cereals will help us address such questions. It may be that differences between wheat and barley domestication are related to the ease with which backcrossing between domesticated and wild cereal plants reintroduces wild traits in generations of barley crops. Ongoing experiments in cereal domestication will provide important information.

Ultimately, our reconstruction of barley domestication, and of its prehistoric importance in human diet and nutrition, depends on interdisciplinary research – the combination of archaeological evidence with botanical, ecological, and experimental evidence. There will always be uncertainties. Archaeological sites have been excavated and analyzed by different researchers practicing different methods and taking plant samples of differing quantity and quality. Modern distributions and plant ecology are the result of historical, environmental, and climatic changes, and ecologists and botanists can only guess in what ways these have affected plant geography. Nevertheless, as more archaeological and botanical evidence emerges, some of these uncertainties may be more conclusively addressed and the process of barley domestication more fully understood.

Joy McCorriston

Endnotes

1. Some of the oldest dates from Jericho can be questioned (Burleigh 1983: 760), and as is the case with wheat remains, domesticated-type barley remains from PPNA Jericho may actually be several hundred years younger than the oldest

- Neolithic radiocarbon dates (10,300 to 10,500 years ago) suggest.
- The earliest unequivocal remains of wheat are the relatively large numbers of domesticated emmer in the lowest levels at Tell Aswad (9,800 years ago), from which no wild wheats were recovered (van Zeist and Bakker-Heeres 1982), and at Jericho, outside the range of wild wheats (Harlan and Zohary 1966).

Bibliography

- Akkermans, P. M. M. G. 1989. The Neolithic of the Balikh Valley, northern Syria: A first assessment. *Paléorient* 15: 122-34.
1991. New radiocarbon dates for the later Neolithic of northern Syria. *Paléorient* 17: 121-5.
- Anderson-Gerfaud, Patricia, Gérard Deraprahamian, and George Willcox. 1991. Les premières cultures de céréales sauvages et domestiques primitives au Proche-Orient Néolithique: Résultats préliminaires d'expérience à Jalès (Ardèche). *Cahiers de l'Euphrate* 5-6: 191-232.
- Bar-Yosef, Ofer, Avi Gopher, Eitan Tchernov, and Mordechai E. Kislev. 1991. Netiv Hagdud: An Early Neolithic village site in the Jordan Valley. *Journal of Field Archaeology* 18: 405-24.
- Bar-Yosef, Ofer, and Mordechai Kislev. 1989. Early farming communities in the Jordan Valley. In *Foraging and farming*, ed. David R. Harris and Gordon C. Hillman, 632-42. London.
- Bar-Yosef, Ofer, and Richard Meadow. 1995. The origins of agriculture in the Near East. In *Last hunters, first farmers*, ed. T. Douglas Price and Anne Birgitte Gebauer, 39-94. Santa Fe, N.Mex.
- Bogucki, Peter. 1996. The spread of early farming in Europe. *American Scientist* 84: 242-53.
- Bothmer, R. von, and N. Jacobsen. 1985. Origin, taxonomy, and related species. In *Barley*, ed. D. C. Rasmussen, 19-56. Madison, Wis.
- Bottero, Jean. 1985. Cuisine of ancient Mesopotamia. *Biblical Archaeologist* 48: 36-47.
- Braidwood, Robert J., and Bruce Howe, eds. 1960. *Prehistoric investigations in Iraqi Kurdistan*. Studies in Ancient Oriental Civilizations, No. 31. Chicago.
- Briggs, D. E. 1978. *Barley*. London.
- Burleigh, Richard. 1983. Additional radiocarbon dates for Jericho. In *Excavations at Jericho*, ed. K. M. Kenyon and T. A. Holland, Vol. 5, 760-5. London.
- Charles, Michael C., and Gordon C. Hillman. 1992. Crop husbandry in a desert environment: Evidence from the charred plant macroremains from Jeitun, 1989-90. In *Jeitun revisited: New archaeological discoveries and paleoenvironmental investigations*, ed. V. M. Masson and David R. Harris (in Russian, manuscript in English). Ashkhabad.
- Childe, V. Gordon. 1951. *Man makes himself*. London.
- Contenson, Henri de. 1985. La région de Damas au Néolithique. *Les annales archéologiques Arabes Syriennes* 35: 9-29.
- Costantini, Lorenzo. 1984. The beginning of agriculture in the Kachi Plain: The evidence of Mehrgarh. In *South Asian archaeology 1981. Proceedings of the 6th international conference of the Association of South Asian Archaeologists in Western Europe*, ed. Bridgit Allchin, 29-33. Cambridge.
- Dietler, Michael. 1990. Driven by drink: The role of drinking in the political economy and the case of Early Iron Age France. *Journal of Anthropological Archaeology* 9: 352-406.
- Edwards, Phillip C. 1988. Natufian settlement in Wadi al-Hammeh. *Paléorient* 14: 309-15.
- Edwards, Phillip C., Stephen J. Bourke, Susan M. Colledge, et al. 1988. Late Pleistocene prehistory in Wadi al-Hammeh, Jordan Valley. In *The prehistory of Jordan: The state of research in 1986*, ed. Andrew N. Garrard and Hans Georg Gebel, 525-65. BAR International Series, No. 396. Oxford.
- Harlan, Jack R. 1968. On the origin of barley. In *Barley: Origin, botany, culture, winterhardiness, genetics, utilization, pests*. U. S. Agricultural Research Service, Agriculture Handbook No. 338, 9-82. Washington, D.C.
1976. Barley. In *Evolution of crop plants*, ed. N. W. Simmonds, 93-8. London.
- Harlan, Jack R., and Daniel Zohary. 1966. Distribution of wild wheats and barley. *Science* 153: 1074-80.
- Harris, David R., V. M. Masson, Y. E. Brezkin, et al. 1993. Investigating early agriculture in central Asia: New research at Jeitun, Turkmenistan. *Antiquity* 67: 324-38.
- Helbaek, Hans. 1960. The paleobotany of the Near East and Europe. In *Prehistoric investigations in Iraqi Kurdistan*, ed. Robert J. Braidwood and Bruce Howe, 99-118. Studies in Ancient Oriental Civilizations, No. 31. Chicago.
- Helbaek, Hans. 1964a. Early Hassunan vegetable [sic] at Es-Sawwan near Sammara. *Sumer* 20: 45-8.
- 1964b. First impressions of the Çatal Hüyük plant husbandry. *Anatolian Studies* 14: 121-3.
1966. Pre-pottery Neolithic farming at Beidha. *Palestine Exploration Quarterly* 98: 61-6.
1969. Plant collecting, dry-farming and irrigation agriculture in prehistoric Deh Luran. In *Prehistory and human ecology of the Deh Luran Plain*, ed. Frank Hole, Kent V. Flannery, and James A. Neeley, 383-426. Memoirs of the Museum of Anthropology, No. 1. Ann Arbor, Mich.
- Hillman, Gordon C. 1975. The plant remains from Tell Abu Hureyra: A preliminary report. In A. M. T. Moore's The excavation of Tell Abu Hureyra in Syria: A preliminary report. *Proceedings of the Prehistoric Society* 41: 70-3.
1989. Late Palaeolithic diet at Wadi Kubbania, Egypt. In *Foraging and farming*, ed. David R. Harris and Gordon C. Hillman, 207-39. London.
1996. Late Pleistocene changes in wild plant-foods available to hunter-gatherers of the northern Fertile Crescent: Possible preludes to cereal cultivation. In *The origin and spread of agriculture and pastoralism in Eurasia*, ed. David R. Harris, 159-203. Washington, D.C.
- Hillman, Gordon C., and M. Stuart Davies. 1990. Measured domestication rates in wild wheats and barley under primitive cultivation, and their archaeological implications. *Journal of World Prehistory* 4: 157-222.
- Hillman, Gordon C., Sue Wales, Frances McClaren, et al. 1993. Identifying problematic remains of ancient plant foods: A comparison of the role of chemical, histological, and morphological criteria. *World Archaeology* 25: 94-121.
- Hopf, Maria. 1983. Jericho plant remains. In *Excavations at Jericho*, ed. Kathleen M. Kenyon and Thomas A. Holland, Vol. 5, 576-621. London.
- Hopf, Maria, and Ofer Bar-Yosef. 1987. Plant remains from

- Hayonim Cave, western Galilee. *Paléorient* 13: 117-20.
- Hough, J. S. 1985. *The biotechnology of malting and brewing*. Cambridge.
- Jarrige, Jean-François, and Richard Meadow. 1980. The antecedents of civilization in the Indus Valley. *Scientific American* 243: 102-10.
- Katz, Solomon, and Fritz Maytag. 1991. Brewing an ancient beer. *Archaeology* 44: 24-33.
- Katz, Solomon, and Mary Voigt. 1986. Bread and beer. *Expedition* 28: 23-34.
- Kislev, Mordechai E. 1988. Dessicated plant remains: An interim report. *Atiqot* 18: 76-81.
1992. Agriculture in the Near East in the seventh millennium B.C. In *Préhistoire de l'agriculture*, ed. Patricia C. Anderson, 87-93. Paris.
- Kislev, Mordechai E., and Ofer Bar-Yosef. 1986. Early Neolithic domesticated and wild barley from Netiv Hagdud region in the Jordan Valley. *Israel Botany Journal* 35: 197-201.
- Kislev, Mordechai E., Dani Nadel, and I. Carmi. 1992. Grain and fruit diet 19,000 years old at Ohalo II, Sea of Galilee, Israel. *Review of Paleobotany and Palynology* 73: 161-6.
- Kozłowski, Stefan K. 1989. Nemrik 9, a PPN Neolithic site in northern Iraq. *Paléorient* 15: 25-31.
- Ladizinsky, Gideon. 1975. Collection of wild cereals in the upper Jordan Valley. *Economic Botany* 29: 264-7.
- Mason, Sarah. 1995. Acorn-eating and ethnographic analogies: A reply to McCorriston. *Antiquity* 69: 1025-30.
- McCorriston, Joy. 1994. Acorn-eating and agricultural origins: California ethnographies as analogies for the ancient Near East. *Antiquity* 68: 97-107.
- Michel, Rudolph H., Patrick E. McGovern, and Virginia R. Badler. 1992. Chemical evidence for ancient beer. *Nature* 360: 24.
- Moore, Jerry D. 1989. Pre-Hispanic beer in coastal Peru: Technology and social contexts of prehistoric production. *American Anthropologist* 91: 682-95.
- Moulins, Dominique de. 1993. Les restes de plantes carbonisées de Çafér Höyük. *Cahiers de l'Euphrate* 7: 191-234.
- Naveh, Zvi. 1974. Effects of fire in the Mediterranean region. In *Fire and ecosystems*, ed. T. T. Kozłowski and C. E. Ahlgren, 401-34. New York.
1984. The vegetation of the Carmel and Sefunim and the evolution of the cultural landscape. In *Sefunim prehistoric sites, Mount Carmel, Israel*, ed. Avraham Ronen, 23-63. BAR International Series, No. 230. Oxford.
- Nesbitt, Mark, and Trevor Watkins. 1995. Collaboration at M'lefaat. In *Qeremez Dere, Tell Afar: Interim Report No. 3*, ed. Trevor Watkins, 11-12. Department of Archaeology, The University of Edinburgh.
- Nevo, Eviatar, Daniel Zohary, A. D. H. Brown, and Michael Haber. 1979. Genetic diversity and environmental associations of wild barley, *Hordeum spontaneum*, in Israel. *Evolution* 33: 815-33.
- Nilan, R. A., and S. E. Ullrich. 1993. Barley: Taxonomy, origin, distribution, production, genetics, and breeding. In *Barley: Chemistry and technology*, ed. Alexander W. MacGregor and Rattan S. Bhatti, 1-29. St. Paul, Minn.
- Noy, Tamar. 1989. Gilgal I - a pre-pottery Neolithic site, Israel - the 1985-1987 seasons. *Paléorient* 15: 11-18.
- Rindos, David R. 1984. *The origins of agriculture*. New York.
- Rosenberg, Michael, R. Mark Nesbitt, Richard W. Redding, and Thomas F. Strasser. 1995. Hallan Çemi Tepesi: Some preliminary observations concerning early Neolithic subsistence behaviors in Eastern Anatolia. *Anatolica* 21: 1-12.
- Stemler, A. B. L., and R. H. Falk. 1980. A scanning electron microscope study of cereal grains from Wadi Kubaniya. In *Loaves and fishes. The prehistory of Wadi Kubaniya*, ed. Fred Wendorf and Angela Close, 299-306. Dallas, Tex.
- Stewart, Robert B. 1976. Paleoethnobotanical report - Çayönü. *Economic Botany* 30: 219-25.
- Stordeur, Danielle. 1989. El Koum 2 Caracol et le PPNB. *Paléorient* 15: 102-10.
- Tchernov, Eitan. 1991. Biological evidence for human sedentism in Southwest Asia during the Natufian. In *The Natufian culture in the Levant*, ed. Ofer Bar-Yosef and François Valla, 315-40. Ann Arbor, Mich.
- van Zeist, Willem. 1972. Palaeobotanical results in the 1970 season at Çayönü, Turkey. *Helinium* 12: 3-19.
- van Zeist, Willem, and Johanna Bakker-Heeres. 1982. Archaeobotanical studies in the Levant. 1. Neolithic sites in the Damascus Basin: Aswad, Ghoraifé, Ramad. *Palaeohistoria* 24: 165-256.
- 1984a. Archaeobotanical studies in the Levant. 3. Late-Palaeolithic Mureybit. *Palaeohistoria* 26: 171-99.
- 1984b. Archaeobotanical studies in the Levant. 2. Neolithic and Halaf levels at Ras Shamra. *Palaeohistoria* 26: 151-98.
- van Zeist, Willem, Phillip E. Smith, R. M. Palfenier-Vegter, et al. 1984. An archaeobotanical study of Ganj Dareh Tepe, Iran. *Palaeohistoria* 26: 201-24.
- van Zeist, Willem, and Willem Waterbolk-van Rooijen. 1985. The palaeobotany of Tell Bouqras, Eastern Syria. *Paléorient* 11: 131-47.
- Watkins, Trevor, Douglas Baird, and Allison Betts. 1989. Qeremez Dere and the Early Aceramic Neolithic of N. Iraq. *Paléorient* 15: 19-24.
- Wendorf, Fred R., Angela E. Close, D. J. Donahue, et al. 1984. New radiocarbon dates on the cereals from Wadi Kubaniya. *Science* 225: 645-6.
- Wendorf, Fred R., Romuald Schild, N. El Hadidi, et al. 1979. The use of barley in the Egyptian Late Palaeolithic. *Science* 205: 1341-7.
- Wetterstrom, Wilma. 1993. Foraging and farming in Egypt: The transition from hunting and gathering to horticulture in the Nile Valley. In *The archaeology of Africa*, ed. Thurston Shaw, Paul Sinclair, Bassey Andah, and Alex Okpoko, 165-226. London.
- Wiebe, G. A. 1968. Introduction of barley into the New World. In *Barley: Origin, botany, culture, winterhardiness, genetics, utilization, pests*. U.S. Department of Agriculture, Agriculture Handbook No. 338, 2-8. Washington, D.C.
- Zohary, Daniel. 1964. Spontaneous brittle six-row barleys, their nature and origin. In *Barley genetics I. Proceedings of the first international barley genetics symposium*, 27-31. Wageningen, Netherlands.
1969. The progenitors of wheat and barley in relation to domestication and agriculture dispersal in the Old World. In *The domestication and exploitation of plants and animals*, ed. Peter J. Ucko and Geoffrey W. Dimbleby, 47-66. London.
1992. Domestication of the Neolithic Near Eastern crop assemblage. In *Préhistoire de l'agriculture*, ed. Patricia C. Anderson, 81-6. Paris.
- Zohary, Daniel, and Maria Hopf. 1993. *Domestication of plants in the Old World*. Second edition. Oxford.

II.A.3. 🌾 Buckwheat

Buckwheat (*Fagopyrum esculentum* Möench) is a crop commonly grown for its black or gray triangular seeds. It can also be grown as a green manure crop, a companion crop, a cover crop, as a source of buckwheat honey (often for the benefit of bees), and as a pharmaceutical plant yielding rutin, which is used in the treatment of capillary fragility. Buckwheat belongs to the *Polygonaceae* family (as do sorrel and rhubarb). Whereas cereals such as wheat, maize, and rice belong to the grass family, buckwheat is not a true cereal. Its grain is a dry fruit.

Buckwheat is believed to be native to Manchuria and Siberia and, reportedly, was cultivated in China by at least 1000 B.C. However, fragments of the grain have been recovered from Japanese sites dating from between 3500 and 5000 B.C., suggesting a much earlier date for the grain's cultivation. It was an important crop in Japan and reached Europe through Turkey and Russia during the fourteenth and fifteenth centuries A.D., although legend would have it entering Europe much earlier with the returning Crusaders. Buckwheat was introduced into North America in the seventeenth century by the Dutch, and it is said that its name derives from the Dutch word *bochweit* (meaning "beech wheat"), because the plant's triangular fruits resemble beechnuts. In German the name for beech is *Buche*, and for buckwheat, *Buchweizen*. Buckwheat has a nutty flavor and, when roasted (*kasha*), a very

strong one. It is a hardy plant that grew in Europe where other grains did not and, thus, supplied peasants in such areas with porridge and pancakes.

Production of buckwheat peaked in the early nineteenth century and has declined since then. During the past decade or so, world production has averaged about 1 million metric tons annually, with the countries of the former Soviet Union accounting for about 90 percent of the total. Other major producing countries are China, Japan, Poland, Canada, Brazil, the United States, South Africa, and Australia. The yield of buckwheat varies considerably by area and by year of production and also with the variety being cultivated. In

Canada, the average yield over the past 10 years has been about 800 kilograms per hectare (kg/ha), although yields of 2,000 kg/ha and higher have been produced.

Types and Cultivars

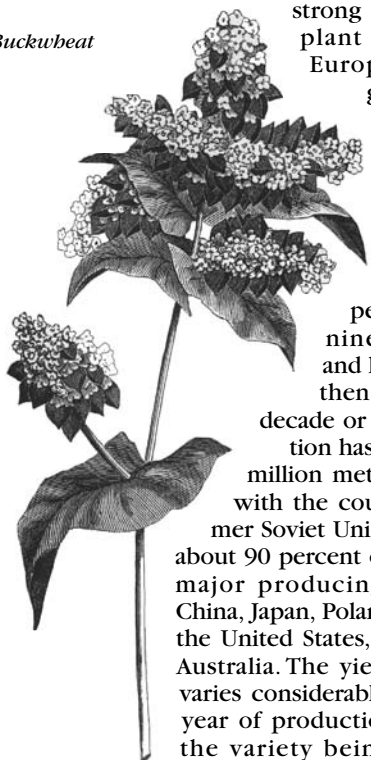
There are three known species of buckwheat: Common buckwheat (*F. esculentum*), tartary buckwheat (*Fagopyrum tataricum*), and perennial buckwheat (*Fagopyrum cymosum*). Common buckwheat is also known as *Fagopyrum sagittatum*, and a form of tartary buckwheat may be called *Fagopyrum kashmirianum*. The cytotaxonomy of buckwheat has not been thoroughly studied, but it is generally believed that perennial buckwheat, particularly the diploid type, is the ancestral form of both tartary buckwheat and common buckwheat.

Tartary buckwheat (also known as rye buckwheat, duck wheat, hull-less, broomless, India wheat, Marino, mountain, Siberian, wild goose, and Calcutta buckwheat) is cultivated in the Himalayan regions of India and China, in eastern Canada, and, occasionally, in mountain areas of the eastern United States. Tartary buckwheat is very frost-resistant. Its seeds – and products made from them – are greenish in color and somewhat bitter in taste. Buckwheat is used primarily as an animal feed or in a mixture of wheat and buckwheat flour. It can also be used as a source of rutin.

Common buckwheat is by far the most economically important species of buckwheat, accounting for over 90 percent of world production. Many types, strains, and cultivars of common buckwheat exist – late-maturing and early-maturing types, Japanese and European types, summer and autumn types. Within a given type there may be strains or varieties with tall or short plants, gray or black seeds, and white or pink flowers. In general, however, common buckwheat varieties from different parts of the world may be divided into two major groups. The first group includes tall, vigorous, late-maturing, photoperiod-sensitive varieties, found in Japan, Korea, southern China, Nepal, and India. Members of the second group are generally insensitive to photoperiod and are small and early-maturing. All of the varieties in Europe and northern China belong to this second group.

Prior to 1950, most producers of buckwheat planted unnamed strains that had been harvested from their own fields or obtained from their neighbors or local stores. Named varieties, developed through plant breeding, were first made available in the 1950s. 'Tokyo', the oldest of the named cultivars introduced into North America, was licensed in 1955 by the Agriculture Canada Research Station in Ottawa. Other cultivars licensed for production in Canada are 'Tempest', 'Mancan', and 'Manor', all developed at the Agriculture Canada Research Station in Morden, Manitoba, since 1965. 'Mancan', which has

Buckwheat



large, dark-brown seeds, thick stems, and large leaves, is the Canadian cultivar preferred in the Japanese market because of its large seeds, desirable flavor and color, and high yield of groats in the milling process.

Cultivars licensed in the United States are 'Pennquad' (released by the Pennsylvania Agricultural Experimental Station and the U.S. Department of Agriculture [USDA] in 1968) and 'Giant American' (a Japanese-type cultivar, apparently developed by a Minnesota farmer). Cultivars developed in the countries of the former Soviet Union since the 1950s include 'Victoria', 'Galleya', 'Eneida', 'Podolyaka', 'Diadema', 'Aelita', and 'Aestoria'. Representative cultivars from other areas of the world include the following: 'Pulawska', 'Emka', and 'Hruszowska' from Poland; 'Bednja 4n' from Yugoslavia; and 'Botan-Soba', 'Shinano No. 1', 'Kyushu-Akisoba Shinshu', and 'Miyazaki Oosoba' from Japan.

Plant and Seed Morphology

Buckwheat is a broad-leaved, erect, herbaceous plant that grows to a height of 0.7 to 1.5 meters. It has a main stem and several branches and can reach full maturity in 60 to 110 days. The stem is usually grooved, succulent, and hollow, except for the nodes. Before maturity, the stems and branches are green to red in color; after maturity, however, they become brown. The plant has a shallow taproot from which branched, lateral roots arise. Its root system is less extensive than those of cereals and constitutes only 3 to 4 percent of the dry weight of the total plant, which - in conjunction with the large leaf surface - may cause wilting during periods of hot and dry weather.

Buckwheat has an indeterminate flowering habit. The flowers of common buckwheat are perfect but incomplete. They have no petals, but the calyx is composed of five petal-like sepals that are usually white, but may also be pink or red. The flowers are arranged in dense clusters at the ends of the branches or on short pedicels arising from the axils of the leaves. Common buckwheat plants bear one of two types of flowers. The pin-type flower has long styles (or female parts) and short stamens (or male parts), and the thrum-type flower has long styles and short pistils. The pistil consists of a one-celled superior ovary and a three-part style with knoblike stigmas and is surrounded by eight stamens. Nectar-secreting glands are located at the base of the ovary. The plants of common buckwheat are generally self-infertile, as self-fertilization is prevented by self-incompatibility of the dimorphic, sporophytic type. Seed production is usually dependent on cross-pollination between the pin and thrum flowers. Honeybees and leaf-cutter bees are effective pollinators that increase seed set and seed yield.

Plants of tartary buckwheat are significantly

different from those of common buckwheat. They have only one flower type and are self-fertile. In addition, they tend to be more husky and more branched and to have narrower, arrow-shaped leaves and smaller, greenish-white flowers. Attempts to transfer the self-compatibility of tartary buckwheat to common buckwheat have proved unsuccessful.

The buckwheat kernel is a triangular, dry fruit (achene), 4 to 9 millimeters (mm) in length, consisting of a hull or pericarp, spermoderm, endosperm, and embryo. Large seeds tend to be concave-sided and small seeds are usually convex-sided. The hull may be glossy, gray, brown, or black and may be solid or mottled. It may be either smooth or rough with lateral furrows. The hulls represent 17 to 26 percent (in tartary buckwheat, 30 to 35 percent) of the kernel weight. Diploid varieties usually have less hull than tetraploids.

Structure of the Kernel

Scanning electron microscopy of the buckwheat kernel has revealed that the hull, spermoderm, endosperm, and embryo are each composed of several layers. For the hull, these are (in order from the outside toward the inside) the epicarp, fiber layers, parenchyma cells, and endocarp. The spermoderm is composed of the outer epiderm, the spongy parenchyma, and the inner epiderm. The endosperm includes an aleurone layer and a subaleurone endosperm, containing starch granules surrounded by a proteinaceous matrix. The embryo, with its two cotyledons, extends through the starchy endosperm. The terminal parts of the cotyledons are often parallel under the kernel surface.

Composition

The gross chemical composition of whole buckwheat grain, the groats, and the hulls is shown in Table II.A.3.1. The mineral and vitamin contents of the whole grains are shown in Table II.A.3.2.

Table II.A.3.1. *Percentage composition of buckwheat seed and its milling products*

Seed or products	Moisture	Ash	Fat	Protein	Fiber	Nitrogen-free extracts
Seed	10.0	1.7	2.4	11.2	10.7	64.0
Groats	10.6	1.8	2.9	11.2	0.9	73.7
Dark flour	11.7	1.2	1.8	8.9	1.0	75.3
Light flour	12.0	1.3	1.6	7.8	0.6	76.7
Very light flour	12.7	0.6	0.5	4.7	0.3	81.1
Hulls	8.0	2.2	0.9	4.5	47.6	36.8
Middlings or shorts	10.7	4.6	7.0	27.2	11.4	39.1

Source: Data from Cole (1931).

Table II.A.3.2. Average mineral and vitamin contents of buckwheat whole grain

Minerals (mg/100 g)		Vitamins (mg/kg)	
Calcium	110	Thiamine	3.3
Iron	4	Riboflavin	10.6
Magnesium	390	Pantothenic acid	11.0
Phosphorus	330	Niacin	18.0
Potassium	450	Pyridoxine	1.5
Copper	0.95	Tocopherol	40.9
Manganese	3.37		
Zinc	0.87		

Source: Data from Marshall and Pomeranz (1982).

Carbohydrates

Starch is quantitatively the major component of buckwheat seed, and concentration varies with the method of extraction and between cultivars. In the whole grain of common buckwheat, the starch content ranges from 59 to 70 percent of the dry matter. The chemical composition of starch isolated from buckwheat grains differs from the composition of cereal starches (Table II.A.3.3). The differences are most pronounced in the case of buckwheat and barley. The amylose content in buckwheat granules varies from 15 to 52 percent, and its degree of polymerization varies from 12 to 45 glucose units. Buckwheat starch granules are irregular, with noticeable flat areas due to compact packing in the endosperm.

Buckwheat grains also contain 0.65 to 0.76 percent reducing sugars, 0.79 to 1.16 percent oligosaccharides, and 0.1 to 0.2 percent nonstarchy polysaccharides. Among the low-molecular-weight sugars, the major component is sucrose. There is also a small amount of arabinose, xylose, glucose, and, probably, the disaccharide melibiose.

Table II.A.3.3. Chemical composition of buckwheat, barley, and corn starch granules smaller than 315 μ

Constituent	Content (% dry matter)		
	Buckwheat	Barley	Corn
Total nitrogen	0.23	0.19	0.12
Free lipids	2.88	2.42	1.80
Ash	0.27	1.04	1.25
Dietary fiber	31.82	41.12	32.02
Hemicellulose ^a	15.58	14.03	24.13
Cellulose ^a	2.72	4.07	0.91
Lignin ^a	13.52	24.03	6.99

^aIn dietary fiber.

Source: Data from Fornal et al. (1987).

Proteins

Protein content in buckwheat varies from 7 to 21 percent, depending on variety and environmental factors during growth. Most currently grown cultivars yield seeds with 11 to 15 percent protein. The major protein fractions are globulins, which represent almost half of all proteins and consist of 12 to 13 subunits with molecular weights between 17,800 and 57,000. Other known buckwheat protein fractions include albumins and prolamins. Reports of the presence of gluten or glutelin in buckwheat seed have recently been discredited.

Buckwheat proteins are particularly rich in the amino acid lysine. They contain less glutamic acid and proline and more arginine, aspartic acid, and tryptophan than do cereal proteins. Because of the high lysine content, buckwheat proteins have a higher biological value (BV) than cereal proteins such as those of wheat, barley, rye, and maize (Table II.A.3.4). Digestibility of buckwheat protein, however, is rather low; this is probably caused by the high-fiber content (17.8 percent) of buckwheat, which may, however, be desirable in some parts of the world. Buckwheat fiber is free of phytic acid and is partially soluble.

Lipids

Whole buckwheat seeds contain 1.5 to 3.7 percent total lipids. The highest concentration is in the embryo (7 to 14 percent) and the lowest in the hull (0.4 to 0.9 percent). However, because the embryo constitutes only 15 to 20 percent of the seed, and the hull is removed prior to milling, the lipid content of the groats is most meaningful. Groats (or dehulled seeds) of 'Mancan', 'Tokyo', and 'Manor' buckwheat contain 2.1 to 2.6 percent total lipids, of which 81 to 85 percent are neutral lipids, 8 to 11 percent phospholipids, and 3 to 5 percent glycolipids. Free lipids, extracted in petroleum ether, range from 2.0 to 2.7 percent. The major fatty acids of buckwheat lipids are palmitic (16:0), oleic (18:1), linoleic (18:2), stearic (18:0), linolenic (18:3), arachidic (20:0), behenic (22:0), and lignoceric (24:0). Of these, the first five are commonly found in all cereals, but the latter three, which represent, on average, 8 percent of the total acids in buckwheat, are only minor components or are not present in cereals.

Phenolic Compounds

The content of phenolics in hulls and groats of common buckwheat is 0.73 and 0.79 percent (and that of tartary buckwheat, 1.87 and 1.52 percent). The three major classes of phenolics are flavonoids, phenolic acids, and condensed tannins. There are many types of flavonoids, three of which are found in buckwheat. These are flavonols, anthocyanins, and C-glycosylflavones. Rutin (quercetin 3-rutinoside), a well-known flavonol diglucoside,

Table II.A.3.4. *Quality of buckwheat and wheat protein*

Parameter	Buckwheat (g/16gN)	Barley (g/16gN)	Wheat (g/16gN)	Rye (g/16gN)	Maize (g/16gN)
Lysine	5.09	3.69	2.55	3.68	2.76
Methionine	1.89	1.82	1.81	1.74	2.37
Cystine	2.02	2.30	1.79	1.99	2.24
Threonine	3.15	3.60	2.84	3.33	3.88
Valine	4.69	5.33	4.50	4.60	5.00
Isoleucine	3.48	3.68	3.39	3.15	3.78
Leucine	6.11	7.11	6.82	5.96	10.51
Phenylalanine	4.19	4.91	4.38	4.41	4.53
Histidine	2.20	2.23	2.30	2.33	2.41
Arginine	8.85	5.38	4.62	5.68	4.35
Tryptophan	1.59	1.11	1.03	1.16	0.62
$N \times 6.25$ (% in dry matter)	12.25	11.42	12.63	10.61	10.06
TD ^a (%)	79.9	84.3	92.4	82.5	93.2
BV ^b (%)	93.1	76.3	62.5	75.4	64.3
NPU ^c (%)	74.4	64.3	57.8	62.2	59.9
UP ^d (%)	9.07	7.34	7.30	6.54	6.03

^aTD = true protein digestibility; ^bBV = biological value; ^cNPU = net protein utilization; ^dUP = utilizable protein = protein $\times$ NPU/10.

Source: Data from Eggum (1980).

used as a drug for the treatment of vascular disorders caused by abnormally fragile or permeable capillaries, occurs in the leaves, stems, flowers, and fruit of buckwheat.

Grading, Handling, and Storage

Grading

In most countries, buckwheat grain is priced according to its physical condition in terms of size, soundness, and general appearance. In Canada, buckwheat is marketed according to grades established under the Canada Grain Act: Grades are No. 1, No. 2, and No. 3 Canada, and Sample. Grade determinants are a minimum test weight of 58 and 55 kilograms per hectoliter (kg/hL) (for Nos. 1 and 2 Canada), variety (designated by size, large or small), degree of sound-

ness, and content of foreign material (Table II.A.3.5). Grades No. 1 and 2 Canada must be free from objectionable odors; No. 3 Canada may have a ground or grassy odor but may not be musty or sour. Test weight, seed size, and foreign material content are determined on a dockage-free sample. Seed size is determined with a No. 8 slotted sieve (3.18 $\times$ 19.05 mm) and becomes part of the grade name (e.g., "buckwheat, No. 1 Canada, large").

"Foreign material" refers to cereal grains (wheat, rye, barley, oats, and triticale), weed seeds, and other grains that are not readily removable by mechanical cleaners, and may include peas, beans, maize, and other domestic or wild weeds. Buckwheat grain containing more than 5 percent foreign material is graded "buckwheat, Sample Canada, (size), account admixture." Damaged seeds include frosted, moldy, distinctly green or otherwise unsound, and dehulled seeds.

Table II.A.3.5. *Primary grade determinants of buckwheat (Canada)*

Grade	kg/hL	Degree of soundness	Maximum limits of foreign material (%)				
			Stones ^a	Ergot	Sclerotinia	Cereal grains	Total foreign material
No. 1	58.0	Well matured, cool and sweet	3	0.0	0.0	1.0%	1.0%
No. 2	55.0	Reasonably well matured, cool and sweet	3	0.05%	0.05%	2.5%	3.0%
No. 3	No min.	May have a ground or grassy odor, but may not be musty or sour	3	0.25%	0.25%	5.0%	5.0%

^aNumber of kernel-size stones in 500 g.

In the United States, buckwheat is not marketed under federally established grades, but some states (for example, Minnesota) have official grain standards that specify the use of Grades 1, 2, 3, and Sample. The grade determinants are similar to those of the Canadian grading system. In Japan, the Buckwheat Millers Association prefers buckwheat that has large, uniform seeds with black hulls and green-colored groats.

Handling

Marketing of buckwheat can be more seriously affected by handling and storage than by other factors such as nutritional quality or processing. The method of handling varies among production areas; nonetheless, in most cases, losses and grain-quality changes occur at postharvest stages. During harvest, in all countries, losses occur, resulting from shattering, germination, depredation by animals, and infection by molds. Threshing is done with combines or by beating the dried plants against stones or wooden bars or by trampling the plants under bullock feet, carts, or tractor wheels.

Transportation of grain from the field to market also results in losses and quality deterioration. Losses during transportation are mainly due to spillage. However, if the grain is exposed to rain or frost during transit, it can subsequently spoil through infection by microorganisms. An efficient system for transportation and distribution of grain must consist of several components, including: (1) collection of grain from farms into consolidated deposits; (2) facilities for short- and long-term storage; (3) loading, unloading, and conveying systems; (4) methods of packaging or bulk handling; (5) roads, railways, and/or waterways; (6) systems for grading the grain and for servicing and maintaining equipment and facilities; (7) systems for recruiting, training, and managing personnel for operation and administration; and (8) systems for research, education, and extension of information to farmers, merchants, and other personnel involved with the overall handling operation.

Storage

Like other grain crops, buckwheat is stored to ensure an even supply over time, to preserve the surplus grain for sale to deficit areas, and for use as seed in the next planting season. Storage of the seeds may be at the farm, trader, market, government, retail, or consumer levels. Storage containers range from sacks to straw huts to bulk storage bins. In developing countries, traditional storage structures include granaries of gunny, cotton, or jute bags as well as those manufactured from reed, bamboo, or wood and plastered with mud and cow dung.

In North America, storage structures include metal, concrete, or wooden bins at the farm level,

Table II.A.3.6. Absorbance of extracted color and tristimulus values of buckwheat samples stored at 25°C and 5 different water activities for 19 months

Water activity	Moisture content (%)	Absorbance index (A ₄₂₀)	Tristimulus values		
			X	Y	Z
0.11	4.1	0.257 b ^a	26.3 c	26.2 c	15.1 b
0.23	6.7	0.233 c	26.7 b	26.5 b	15.1 b
0.31	8.7	0.241 c	27.1 a	27.0 a	15.3 a
0.51	13.0	0.290 a	25.3 d	24.7 d	13.4 d
0.67	13.8	0.238 c	26.4 bc	25.9 c	14.2 c

^aMeans separated by Duncan's multiple range test, 0.01 level of probability.

Source: Data from Mazza (1986).

elevators and annexes at centralized receiving, storage, and shipping points, and concrete silos at grain terminals.

Bagged buckwheat is highly susceptible to attack by insects and rodents. Hence, bulk storage in bins, elevators, and silos is best. Grain bins made of wood are usually square and, by virtue of their construction, possess a multitude of cracks, crevices, and angles that are havens for insects and their eggs and larvae. Concrete bins are usually round, star-shaped, or hexagonal, and star-shaped bins also have crevices that can harbor grain residues, constituting a source of infestation. Moreover, concrete possesses certain sorptive properties and chemical reactivity, and unless coated with an impervious material such as paint, the walls of concrete bins can interfere with fumigation procedures. Metal bins are usually round, possess few crevices, and do not react significantly with protective chemicals.

Neither concrete nor metal allows interchanges between stored grain and the atmosphere; moisture movement resulting from temperature fluctuations, convection, and condensation can result in deterioration and even internal combustion of the grains. A moisture content of 16 percent or less is required for the safe storage of buckwheat. If the seed requires drying, the temperature of the drying air should not exceed 43°C.

During storage at ambient temperature and relative humidity, the color of the aleurone layer changes from a desirable light green to the undesirable reddish brown. This undesirable quality change can be reduced by storing the seed at a lower temperature and at a relative humidity below 45 percent. Table II.A.3.6 gives the absorbance of the extracted color of buckwheat samples stored at 25°C and 0.11 to 0.67 water activity for 19 months. Maximum browning-pigment production occurs at 0.45 to 0.55 water activity, or 45 to 55 percent relative humidity.

Primary Processing

Primary processing of buckwheat includes cleaning, dehulling, and milling. The aim of seed cleaning is to remove other plant parts, soil, stones, weed seeds, chaff, dust, seeds of other crops, metallic particles, and small and immature buckwheat seeds. The extent and sophistication of the cleaning equipment depends largely on the size of the operation and the requirements for the finished product(s). Milling of buckwheat seed can be carried out by virtually any equipment capable of milling cereal grains. Hammer mills, stone mills, pin mills, disk mills, and roller mills have all been used to mill buckwheat. Of these, stone mills and roller mills are probably the most extensively used today.

The milling process may be of two types. In the first and most common type, the whole seeds are first dehulled and then milled. In the second type, the seeds are milled and then screened to remove the hulls. When dehulling and milling are separate operations, the seeds are segregated according to size and may be steamed and dried prior to dehulling. The latter procedure is carried out by impact or abrasion against emery stones or steel, followed by air- or screen-separation of groats and hulls. A widely used buckwheat dehuller is built on the principle of stone-milling, with emery stones set to crack the hull without breaking the groat. The effectiveness of this type of dehuller depends on the clearance between the seed cracking surfaces, and for any seed size there is an optimal setting. The ease of dehulling and the percentage of recovery of

undamaged groats depends on variety and moisture content (Table II.A.3.7). From the dehuller, the groats go over sieves of different mesh for sizing into whole groats and two or more sizes of broken groats. Flour is produced by passing the groats through stone and/or roller grinders.

When buckwheat seed is to be processed into flour only, and production of groats is not a requirement, the seeds are ground on break rolls or stone mills and then screened to separate the coarse flour from the hulls. The coarse flour is further reduced by a series of size reduction rolls, each grinding operation followed by a sifting to fractionate the mixture of particles according to their size (Figure II.A.3.1). The flour yield ranges from 50 to 75 percent depending on the size, shape, and condition of the seeds and the efficiency of the dehulling and milling operations.

End Products

Buckwheat flour is generally dark in color because of the presence of hull fragments. In North America, it is used primarily for making buckwheat pancakes and is commonly marketed in the form of prepared mixes. These mixes generally contain buckwheat flour mixed with wheat, maize, rice, oat, or soybean flours and a leavening agent. Buckwheat is also used with vegetables and spices in kasha and soup mixes, and with wheat, maize, or rice in ready-to-eat breakfast products, porridge, bread, and pasta products.

Table II.A.3.7. *Influence of cultivar and moisture content on debulling characteristics and color of buckwheat seeds stored at 25°C and water activities of 0.23–0.97 for 45 days*

Cultivar	Water activity	Moisture content (%)	Dehulling recovery (%)	Dehulled groat		Hunter color values ^a					
				Whole (%)	Broken (%)	Whole groat		Broken groat		Mixed groat	
						L	a	L	a	L	a
Mancan	0.23	5.98	69.2	30.1	69.9	52.0	+0.1	68.7	+0.3	63.5	+0.1
	0.52	9.79	67.0	33.3	66.7	53.0	−0.1	60.0	+0.4	58.7	+0.5
	0.75	13.47	65.4	44.5	55.5	52.3	+0.3	56.5	−0.4	57.4	+0.4
	0.97	19.80	65.0	68.5	31.5	51.0	+0.4	57.87	+1.3	53.2	+0.7
Tokyo	0.23	5.84	66.3	28.7	71.3	51.7	+0.7	62.7	+0.5	60.6	+0.3
	0.52	9.77	60.5	33.5	66.5	51.1	+0.7	58.4	+0.1	57.6	+1.3
	0.75	13.30	58.2	45.5	54.5	51.3	+0.7	54.8	+0.3	54.6	+0.8
	0.97	18.74	51.6	75.1	24.9	50.3	+1.8	58.3	+1.1	52.1	+1.0
Manor	0.23	5.93	54.9	37.1	62.9	52.6	+1.4	60.6	+0.8	59.0	+1.0
	0.52	9.90	50.4	35.7	64.3	52.9	+1.9	58.0	+0.7	58.9	−0.7
	0.75	13.50	41.6	48.4	51.6	52.9	+1.5	56.6	+1.4	54.5	+1.2
	0.97	19.13	32.5	61.5	38.5	51.7	+2.1	61.3	+1.7	53.2	+0.4

^aL = lightness; a = redness when positive and greenness when negative.

Source: Data from Mazza and Campbell (1985).

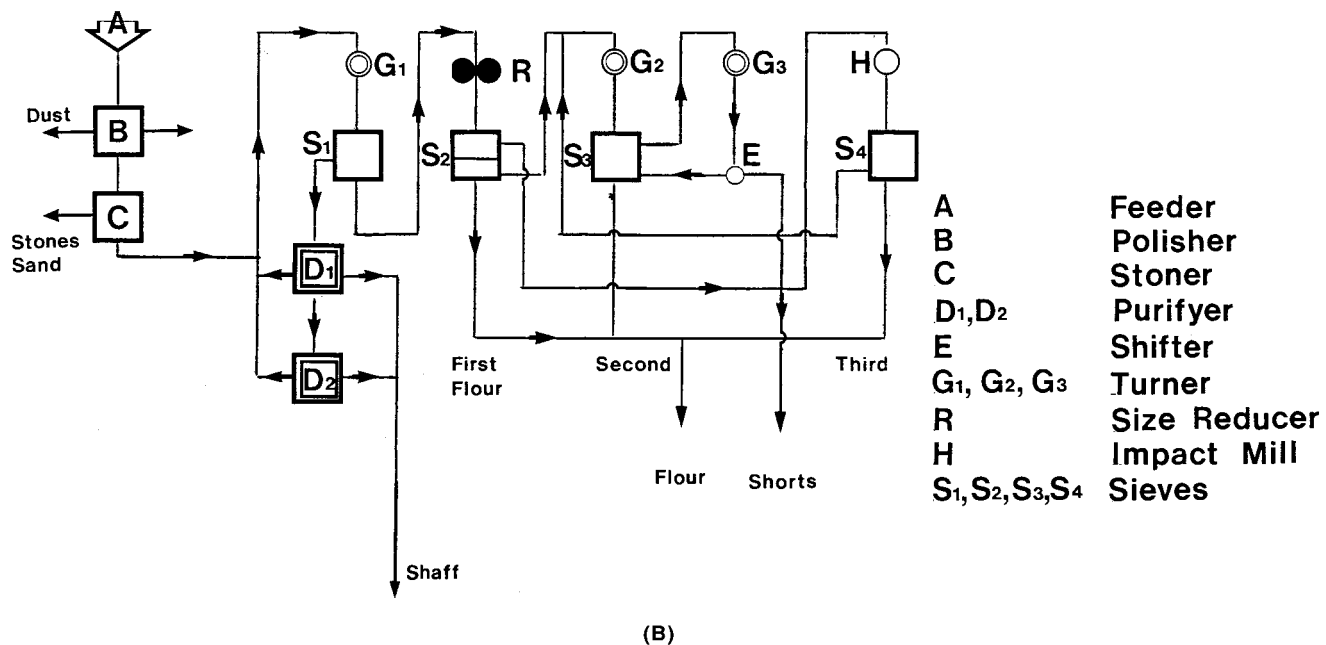
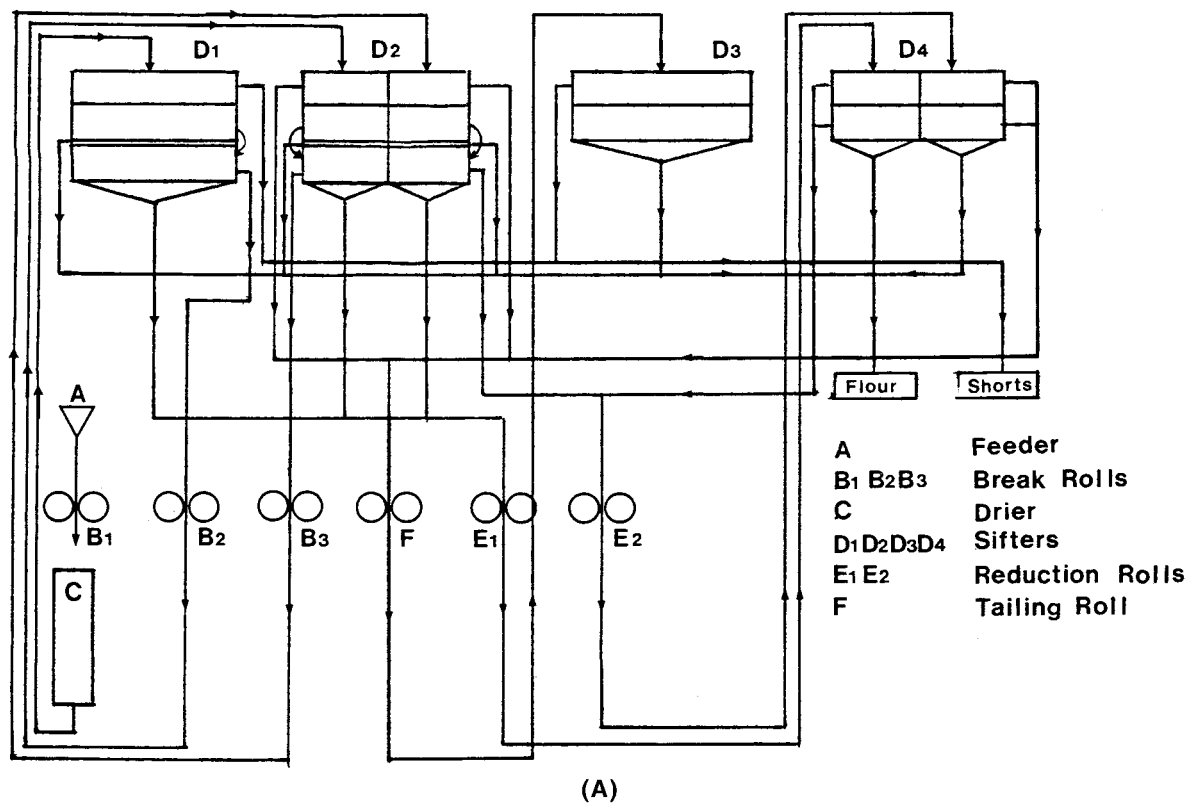


Figure II.A.3.1. Flow diagram of two buckwheat mills: (A) roller mill; (B) stone-roller mill.

In Japan, buckwheat flour is used primarily for making *soba* or *sobakiri* (buckwheat noodles) and *Teuchi Soba* (handmade buckwheat noodles). These products are prepared at *soba* shops or at home from a mixture of buckwheat and wheat flours. The wheat flour is used because of its binding properties and availability. *Soba* is made by hand or mechanically. In both methods, buckwheat and wheat flours are mixed with each other and then with water to form a stiff dough that is kneaded, rolled into a thin sheet (1.4 mm) with a rolling pin or by passing it between sheeting rolls, and cut into long strips. The product may be cooked immediately, sold fresh, or dried. For consumption, the noodles are boiled in hot water, put into bamboo baskets, and then dipped into cold water.

In Europe, most buckwheat is milled into groats that are used in porridge, in meat products (especially hamburger), or consumed with fresh or sour milk. A mixture of buckwheat groats with cottage cheese, sugar, peppermint, and eggs is employed as stuffing in a variety of dumplings. Buckwheat flour is used with wheat or rye flour and yeast to make fried specialty products such as bread, biscuits, and other confectioneries. An extended ready-to-eat breakfast product of high nutritional value, made from maize and buckwheat, is produced and marketed in Western Europe. This product contains over 14 percent protein and 8 percent soluble fiber. Similar products have also been developed in Poland and the former Soviet Union. In most countries, the quality of buckwheat end products is controlled by law.

The pace of development of new food products from buckwheat is expected to increase. This will likely parallel the increasing consumer demand for foods capable of preventing or alleviating disease and promoting health.

G. Mazza

Bibliography

- Campbell, C. G., and G. H. Gubbels. 1986. Growing buckwheat. *Agriculture Canada Research Branch Technical Bulletin 1986-7E*. Agriculture Canada Research Station, Morden, Manitoba.
- Cole, W. R. 1931. Buckwheat milling and its by-products. United States Department of Agriculture, Circular 190. Washington, D.C.
- DeJong, H. 1972. Buckwheat. *Field Crops Abstracts* 25: 389-96.
- Eggum, B. O. 1980. The protein quality of buckwheat in comparison with other protein sources of plant or animal origin. Buckwheat Symposium, Ljubljana, Yugoslavia, September 1-3, 115-20.
- Fornal, L., M. Soral-Smietana, Z. Smietana, and J. Szpendowski. 1987. Chemical characteristics and physico-chemical properties of the extruded mixtures of cereal starches. *Stärke* 39: 75-8.
- Institute of Soil Science and Plant Cultivation, ed. 1986. *Buckwheat research 1986*. Pulawy, Poland.
- Kreft, I., B. Javornik, and B. Dolisek, eds. 1980. Buckwheat genetics, plant breeding and utilization. VTOZD za agronomijo Biotech. Ljubljana, Yugoslavia.
- Marshall, H. G., and Y. Pomeranz. 1982. Buckwheat: Description, breeding, production and utilization. In *Cereals '78: Better nutrition for the world's millions*, ed. Y. Pomeranz, 201-17. St. Paul, Minn.
- Mazza, G. 1986. Buckwheat browning and color assessment. *Cereal Chemistry* 63: 362-6.
- Mazza, G., and C. G. Campbell. 1985. Influence of water activity and temperature on dehulling of buckwheat. *Cereal Chemistry* 62: 31-4.
- Nagatoma, T., and T. Adachi, eds. 1983. *Buckwheat research 1983*. Miyazaki, Japan.
- Oomah, B. D., and G. Mazza. 1996. Flavonoids and antioxidative activities in buckwheat. *Journal of Agricultural and Food Chemistry* 44: 1746-50.

II.A.4. 🌽 Maize

Maize (*Zea mays* L.), a member of the grass family Poaceae (synonym Gramineae), is the most important human dietary cereal grain in Latin America and Africa and the second most abundant cultivated cereal worldwide. Originating in varying altitudes and climates in the Americas, where it still exhibits its greatest diversity of types, maize was introduced across temperate Europe and in Asia and Africa during the sixteenth and seventeenth centuries.

Corn



It became a staple food of Central Europe, a cheap means of provisioning the African-American slave trade by the end of the eighteenth century, and the usual ration of workers in British mines in Africa by the end of the nineteenth century. In the twentieth century, major increases in maize production, attributed to developments in maize breeding, associated water management, fertilizer response, pest control, and ever-expanding nutritional and industrial uses, have contributed to its advance as an intercrop (and sometimes as a staple) in parts of Asia and to the doubling and tripling of maize harvests throughout North America and Europe. High-yield varieties and government agricultural support and marketing programs, as well as maize's biological advantages of high energy yields, high extraction rate, and greater adaptability relative to wheat or rice, have all led to maize displacing sorghum and other grains over much of Africa.

On all continents, maize has been fitted into a wide variety of environments and culinary preparations; even more significant, however, it has become a component of mixed maize-livestock economies and diets. Of the three major cereal grains (wheat, rice, and maize), maize is the only one not grown primarily for direct human consumption. Approximately one-fifth of all maize grown worldwide is eaten directly by people; two-thirds is eaten by their animals; and approximately one-tenth is used as a raw material in manufactured goods, including many non-food products.

Maize Literature

Principal sources for understanding the diverse maize cultures and agricultures are P. Weatherwax's (1954) *Indian Corn in Old America*, an account of sixteenth-century maize-based agriculture and household arts; S. Johannessen and C. A. Hastorf's (1994) *Corn and Culture*, essays that capture New World archaeological and ethnographic perspectives; and H. A. Wallace and E. N. Bressman's (1923) *Corn and Corn Growing*, and Wallace and W. L. Brown's (1956) *Corn and Its Early Fathers*, both of which chronicle the early history of maize breeding and agribusiness in the United States. The diffusion of maize in the Old World has been traced in J. Finan's (1950) summary of discussions of maize in fifteenth- and sixteenth-century herbals, in M. Bonafous's (1836) *Natural Agricultural and Economic History of Maize*, in A. de Candolle's (1884) *Origin of Cultivated Plants*, and in D. Roe's (1973) *A Plague of Corn: The Social History of Pellagra*. B. Fussell's (1992) *The Story of Corn* applies the art of storytelling to maize culinary history, with special emphasis on the New World. Quincentenary writings, celebrating America's first cuisines and the cultural and nutritional influence of maize, as well as that of other New World indigenous crops, include works by W. C. Galinat (1992), S. Coe (1994), and J. Long (1996).

More recent regional, cultural, agricultural, and economic perspectives highlight the plight of Mexico's peasant farmers under conditions of technological and economic change (Hewitt de Alcantara 1976, 1992; Montanez and Warman 1985; Austin and Esteva 1987; Barkin, Batt, and DeWalt 1990), the displacement of other crops by maize in Africa (Miracle 1966), and the significance of maize in African "green revolutions" (Eicher 1995; Smale 1995). *The Corn Economy of Indonesia* (Timmer 1987) and C. Dowsell, R. L. Paliwal, and R. P. Cantrell's (1996) *Maize in the Third World* explore maize's growing dietary and economic significance in developing countries. The latter includes detailed country studies of Ghana, Zimbabwe, Thailand, China, Guatemala, and Brazil.

Global developments in maize breeding, agronomy, and extension are chronicled in the publications of Mexico's International Center for the Improvement of Maize and Wheat (CIMMYT), especially in *World Maize Facts and Trends* (CIMMYT 1981, 1984, 1987, 1990, 1992, 1994), and in research reports and proceedings of regional workshops. Maize genetics is summarized by David B. Walden (1978), G. E. Sprague and J. W. Dudley (1988), and the National Corn Growers Association (1992). Molecular biologists who use maize as a model system share techniques in the *Maize Genetics Cooperation Newsletter* and *The Maize Handbook* (Freeling and Walbot 1993). One explanation for the extensive geographic and cultural range of maize lies in its unusually active "promoter," or "jumping," genes and extremely large chromosomes, which have made it a model plant for the study of genetics - the *Drosophila* of the plant world.

Geographic Range

Maize is grown from 50 degrees north latitude in Canada and Russia to almost 50 degrees south latitude in South America, at altitudes from below sea level in the Caspian plain to above 12,000 feet in the Peruvian Andes, in rainfall regions with less than 10 inches in Russia to more than 400 inches on Colombia's Pacific coast, and in growing seasons ranging from 3 to 13 months (FAO 1953). Early-maturing, cold-tolerant varieties allow maize to penetrate the higher latitudes of Europe and China, and aluminum-tolerant varieties increase production in the Brazilian savanna. In the tropics and subtropics of Latin America and Asia, maize is double- or triple-cropped, sometimes planted in rotation or "relay-cropped" with wheat, rice, and occasionally, soybeans, whereas in temperate regions it is monocropped, or multi-cropped with legumes, cucurbits, and roots or tubers. In North America, it is planted in rotation with soybeans.

Yields average 2.5 tons per hectare in developing countries, where maize is more often a component of less input-intensive multicrop systems, and 6.2

tons per hectare in industrialized countries, where maize tends to be input-intensive and single-cropped. The U.S. Midwest, which produced more than half of the total world supply of maize in the early 1950s, continued to dominate production in the early 1990s, with 210.7 million tons, followed by China (97.2 million tons), and then Brazil (25.2 million tons), Mexico (14.6 million tons), France (12.2 million tons), India (9.2 million tons), and the countries of the former Soviet Union (9.0 million tons). Developing countries overall account for 64 percent of maize area and 43 percent of world harvests (FAO 1993). The United States, China, France, Argentina, Hungary, and Thailand together account for 95 percent of the world maize trade, which fluctuates between 60 and 70 million tons, most of which goes into animal feed.

Cultural Range

Maize serves predominantly as direct human food in its traditional heartlands of Mexico, Central America, the Caribbean, and the South American Andes, as well as in southern and eastern Africa, where the crop has replaced sorghum, millet, and sometimes roots and tuber crops in the twentieth century. The highest annual per capita intakes (close to 100 kilograms per capita per year) are reported for Mexico, Guatemala, and Honduras, where the staple food is tortillas, and for Kenya, Malawi, Zambia, and Zimbabwe, where the staple is a porridge. Maize is also an essential regional and seasonal staple in Indonesia and parts of China.

However, maize is considerably more significant in the human food chain when it first feeds livestock animals that, in turn, convert the grain into meat and dairy products. In the United States, 150 million tons of maize were harvested for feed in 1991; in Germany, three-fourths of the maize crop went for silage. In some developing countries, such as Pakistan, India, and Egypt, the value of maize fodder for bovines may surpass that of the grain for humans and other animals. In Mexico, the "Green Revolution" Puebla Project developed improved, tall (versus short, stiff-strawed) varieties in response to demands for fodder as well as grain.

Since World War II, processing for specialized food and nonfood uses has elevated and diversified maize's economic and nutritional significance. In the United States, for example, maize-based starch and sweeteners account for 20 million tons (15 million tons go to beverages alone), cereal products claim 3 million tons, and distilled products 0.3 million tons. Maize-based ethanol, used as a fuel extender, requires 10 million tons; and plastics and other industrial products also employ maize. The geographic and cultural ranges of maize are tributes to its high mutation rate, genetic diversity and adaptability, and continuing cultural selection for desirable characteristics.

Biology and Biodiversity

More than 300 races of maize, consisting of hundreds of lineages and thousands of cultivars, have been described. But the precise ancestry of maize remains a mystery, and geographical origins and distributions are controversial.

Biological Evolution

Teosinte (*Zea* spp.), a weedy grass that grows in Mexico and Guatemala, and *Tripsacum*, a more distantly related rhizomatous perennial, are maize's closest wild relatives. All three species differ from other grasses in that they bear separate male and female flowers on the same plant. Key morphological traits distinguishing maize from its wild relatives are its many-rowed ear compared to a single-rowed spike, a rigid rather than easily shattered rachis, a pair of kernels in each cupule compared to a single grain per cupule, and an unprotected or naked grain compared to seed enclosed in a hard fruitcase. Unlike the inflorescence structures of other grasses, maize produces a multirowed ear with hundreds of kernels attached to a cob that is enclosed by husks, which makes it amenable for easy harvest, drying, and storage.

Based on interpretations of evidence from cytology, anatomy, morphology, systematics, classical and molecular genetics, experimental breeding, and archaeology, there are three recognized theories about the origins of maize: (1) the ancestor of maize is annual teosinte (Vavilov 1931; Beadle 1939; Iltis 1983; Kato 1984); (2) maize evolved from an as yet undiscovered wild maize or other ancestor (Mangelsdorf 1974); and (3) maize derived from hybridization between teosinte and another wild grass (Harshberger 1896; Eubanks 1995). Although the most popular theory holds that teosinte is the progenitor, present evidence does not clearly resolve its genetic role. Firmer evidence supports the idea that introgression of teosinte germ plasm contributed to the rapid evolution of diverse maize land races in prehistory (Wellhausen et al. 1952). Teosintes—which include two annual subspecies from Mexico (*Z. mays* ssp. *mexicana* and ssp. *parviglumis*), two annual species from Guatemala (*Z. huebuetenangensis* and *Z. luxurians*), and two perennial species from Mexico (*Z. perennis* and *Z. diploperennis*)—have the same base chromosome number ($n = 10$) as maize and can hybridize naturally with it. Like maize, teosintes bear their male flowers in tassels at the summit of their main stems and their female flowers laterally in leaf axils. Although the ears of teosinte and maize are dramatically different, teosinte in architecture closely mimics maize before flowering, and—so far—no one has demonstrated effectively how the female spike might have been transformed into the complex structure of a maize ear.

Tripsacum spp. have a base chromosome number of $n = 18$ and ploidy levels ranging from $2n = 36$ to

$2n = 108$. *Tripsacum* is distinctive from maize and teosinte because it bears male and female flowers on the same spike, with the staminate (male) flowers directly above the pistillate (female) flowers. This primitive trait is seen in some of the earliest prehistoric maize (on Zapotec urns from the Valley of Oaxaca, Mexico, c. A.D. 500–900, which depict maize with staminate tips) and also in some South American races. *Tripsacum* plants also frequently bear pairs of kernels in a single cupule, another maize trait. The ears of F1 *Tripsacum*-teosinte hybrids have pairs of exposed kernels in fused cupules and resemble the oldest archaeological maize remains from Tehuacan, Mexico (Eubanks 1995). Although the theory that domesticated maize arose from hybridization between an unknown wild maize and *Tripsacum* is no longer accepted, and crosses between *Tripsacum* and maize or annual teosinte are almost always sterile, crosses between *Tripsacum* and perennial teosinte have been shown to produce fully fertile hybrid plants (Eubanks 1995), and *Tripsacum* has potential as a source of beneficial traits for maize improvement.

Molecular evidence for maize evolution includes analyses of isozymes and DNA of nuclear and cytoplasmic genes. Results indicate that isozyme analysis cannot fully characterize genetic variation in *Zea*, and application of this technique to understanding evolutionary history is limited. In addition, certain maize teosintes (*Z. m. parviglumis* and *Z. m. mexicana*), thought to be ancestral to maize, may actually post-date its origin. In sum, the origins of maize remain obscure.

Geographic Origin and Distribution

Most scientists concur that maize appeared 7,000 to 10,000 years ago in Mesoamerica, but controversy surrounds whether maize was domesticated one or more times and in one or more locations. Based on racial diversity and the presence of teosinte in Mexico but not Peru, N. I. Vavilov (1931) considered Mexico to be the primary center of origin. The earliest accepted archaeological evidence comes from cave deposits in Tehuacan, Puebla, in central Mexico. The cobs found there ranged from 19 to 25 millimeters (mm) long and had four to eight rows of kernels surrounded by very long glumes. The remarkably well-preserved specimens provide a complete evolutionary sequence of maize dating from at least as far back as 3600 B.C. up to A.D. 1500. Over this time, tiny eight-rowed ears were transformed into early cultivated maize and then into early tripsacoid maize, ultimately changing into the Nal Tel-Chapalote complex, late tripsacoid, and slender popcorn of later phases (Mangelsdorf 1974). An explosive period of variation, brought about by the hybridization of maize with teosinte, began around 1500 B.C. (Wilkes 1989).

From Mexico, maize is thought to have moved south and north, reaching Peru around 3000 B.C. and

North America sometime later. However, pollen identified as maize was present with phytoliths in preceramic contexts in deposits dated to 4900 B.C. in Panama and in sediments dated to 4000 B.C. in Amazonian Ecuador. Although the identification of maize pollen and phytoliths (as opposed to those of a wild relative) remains uncertain, some investigators (Bonavia and Grobman 1989) have argued that such evidence, combined with maize germ plasm data, indicates the existence of a second center of domestication in the Central Andean region of South America, which generated its own distinct racial complexes of the plant between 6000 and 4000 B.C. Fully developed maize appeared later in the lowlands (Sanoja 1989).

Maize arrived in North America indisputably later. Flint varieties adapted to the shorter nights and frost-free growing seasons of the upper Midwest evolved only around A.D. 1100, although maize had been introduced at least 500 years earlier. Ridged fields later allowed cultivators to expand the growing season by raising soil and air temperatures and controlling moisture. In the Eastern Woodlands, 12-row (from A.D. 200 to 600) and 8-row (from around 800) varieties supplemented the existing starchy-seed food complexes (Gallagher 1989; Watson 1989).

Germ Plasm and Genetic Diversity

Gene banks have collected and now maintain 90 to 95 percent of all known genetic diversity of maize. The largest collections are held by the Vavilov Institute (Russia) and the Maize Research Institute (Yugoslavia), which contain mostly Russian and European accessions. The genetically most diverse New World collections are maintained at the National Seed Storage Laboratory in the United States, CIMMYT and the Instituto Nacional de Investigaciones Forestales y Agropecuarias (INIFAP – the National Institute of Forestry and Agricultural Research) in Mexico, the National Agricultural University in Peru, the National Agricultural Research Institute in Colombia, the Brazilian Corporation of Agricultural Research (EMBRAPA), the Instituto de Nutrición y Tecnología de los Alimentos (INTA – the Institute of Nutrition and Food Technology) at the University of Chile, Santiago, and the National Agricultural Research Institute (INIA) in Chile. International maize breeding programs operate at CIMMYT and at the International Institute for Tropical Agriculture (IITA), which interface with national maize breeding programs in developing countries (Dowswell, Paliwal, and Cantrell 1996).

The germ plasm collections begun by the Rockefeller Foundation and the Mexican Ministry of Agriculture in 1943 (which classified maize according to productiveness, disease resistance, and other agronomic characteristics) have since been supplemented by the collections of international agricultural research centers that treat additional genetic, cytological, and botanical characteristics (Wellhausen et al. 1952;

Mangelsdorf 1974). All contribute information for the contemporary and future classification and breeding of maize.

Maize Classifications

Maize plants range from 2 to 20 feet in height, with 8 to 48 leaves, 1 to 15 stalks from a single seed, and ears that range from thumb-sized (popcorn) to 2 feet in length. The different varieties have different geographical, climatic, and pest tolerances. The mature kernel consists of the pericarp (thin shell), endosperm (storage organ), and embryo or germ, which contains most of the fat, vitamins, and minerals and varies in chemical composition, shape, and color.

The principal maize classifications are based on grain starch and appearance – these characteristics influence suitability for end uses. In “flints,” the starch is hard. In “dents,” the kernel is softer, with a larger proportion of floury endosperm and hard starch confined to the side of the kernel. “Floury” varieties have soft and mealy starch; “pop” corns are very hard. “Sweet” corns have more sugar, and “waxy” maizes contain starch composed entirely of amylopectin, without the 22 percent amylose characteristic of dents.

Dents account for 95 percent of all maize. The kernels acquire the characteristic “dent” when the grain is dried and the soft, starchy amylose of the core and the cap contract. Most dent maize is yellow and is fed to livestock; white dents are preferred for human food in Mexico, Central America, the Caribbean, and southern Africa. Flint maize, with its hard outer layer of starch, makes a very good-quality maize meal when dry-milled. It stores more durably than other types because it absorbs less moisture and is more resistant to fungi and insects. Flints germinate better in colder, wetter soils, mature earlier, and tend to perform well at higher latitudes. Popcorns are extremely hard flint varieties; when heated, the water in the starch steam-pressures the endosperm to explode, and the small kernels swell and burst.

Sweet corns are varieties bred especially for consumption in an immature state. A number of varieties of sweet corn, exhibiting simple mutations, were developed as garden vegetables in the United States beginning around 1800. Sweet varieties known as *sara chulpi* were known much earlier in the Andes, where they were usually parched before eating. Floury maizes are grown in the Andean highlands of South America, where they have been selected for beer making and special food preparations (*kancha*), and in the U.S. Southwest, where they are preferred for their soft starch, which grinds easily. Waxy varieties are grown for particular dishes in parts of Asia and for use in industrial starches in the United States.

In addition, maize grains are classified by color, which comes mostly from the endosperm but is also influenced by pigments in the outer aleurone cell

layer and pericarp. Throughout most of the world, white maize is preferred for human consumption and yellow for animal feed, although Central and Mediterranean Europeans eat yellow maize and indigenous Americans carefully select blue (purple, black), red, and speckled varieties for special regional culinary or ritual uses. Color is probably the most important classification criterion among New World indigenous cultivators, who use color terms to code information on the ecological tolerances, textures, and cooking characteristics of local varieties.

Breeding

The early indigenous cultivators of maize created “one of the most heterogeneous cultivated plants in existence” (Weatherwax 1954: 182). They selected and saved seed based on ear form, row number, and arrangement; kernel size, form, color, taste, texture, and processing characteristics; and plant-growth characteristics such as size, earliness, yield, disease resistance, and drought tolerance. Traditional farmers planted multiple varieties as a hedge against stressors, and Native American populations have continued this practice in the United States (Ford 1994) and Latin America (Brush, Bellon, and Schmidt 1988; Bellon 1991). However, only a small fraction of the biodiversity of traditional maize was transported to North America, to Europe, from Europe to Asia and Africa, and back to North America.

During all but the last hundred years, maize breeding involved open-pollinated varieties – varieties bred true from parent to offspring – so that farmers could select, save, and plant seed of desirable maize types. Hybrid maize, by contrast, involves crossing two inbred varieties to produce an offspring that demonstrates “hybrid vigor,” or heterosis (with yields higher than either parent). But the seed from the hybrid plant will not breed true. Instead, new generations of hybrid seed must be produced in each succeeding generation through controlled crosses of the inbred lines. Consequently, save for producing their own crosses, farmers must purchase seed each cultivation season, which has given rise to a large hybrid seed industry, particularly in developed countries.

Hybrid maize had its beginnings in the United States in 1856 with the development by an Illinois farmer of Reid Yellow Dent, a mixture of Southern Dent and Northern Flint types that proved to be high-yielding and resistant to disease. There followed a series of scientific studies demonstrating that increased yields (hybrid vigor) resulted from the crossing of two inbred varieties. W. J. Beal, an agrobotanist at Michigan State University, in 1877 made the first controlled crosses of maize that demonstrated increased yields. Botanist George Shull, of Cold Spring Harbor, New York, developed the technique of inbreeding. He showed that although self-pollinated plants weakened generation after genera-

tion, single crosses of inbred lines demonstrated heterosis, or hybrid vigor. Edward East, working at the Connecticut Agricultural Experimental Station during the same period, developed single-cross inbred hybrids with 25 to 30 percent higher yields than the best open-pollinated varieties. A student of East's, D. F. Jones, working with Paul Mangelsdorf, in 1918 developed double-cross hybrids, which used two single-cross hybrids rather than inbred lines as parents and overcame the poor seed-yields and weakness of inbred lines so that hybrid seeds became economically feasible.

By the late 1920s, private seed companies were forming to sell high-yield hybrid lines. Henry A. Wallace, later secretary of agriculture under U.S. president Franklin Roosevelt, established Pioneer Hi-Bred for the production and sale of hybrid seed in 1926, in what he hoped would herald a new era of productivity and private enterprise for American agriculture. From the 1930s through the 1950s, commercial hybrids helped U.S. maize yields to increase, on average, 2.7 percent per year. By the mid-1940s, hybrids covered almost all of the U.S. "Corn Belt," and advances were under way in hybrid seeds that were adapted to European, Latin American, and African growing conditions. Another quantum leap in yields was achieved after World War II through chemical and management techniques. New double- and triple-cross hybrids responsive to applications of nitrogen fertilizers substantially raised yields in the 1960s, and again in the 1970s, with the release of a new generation of fertilizer-responsive single-cross hybrids that were planted in denser stands and protected by increased quantities of pesticides. In the 1980s, however, concerns about cost reduction, improved input efficiency, and natural resource (including biodiversity) conservation supplanted the earlier emphasis on simply increasing yields, with the result that yields remained flat.

Breeders also have been concerned with diversifying the parent stock of inbred hybrids, which are formed by the repeated self-pollination of individual plants and which, over generations, become genetically uniform and different from other lines. To prevent self-pollination when two inbred lines are crossed to produce a hybrid, the tassels are removed from the male parent. Discovery of lines with cytoplasmic male sterility allowed this labor-intensive step to be eliminated, but although it was desirable for the seed industry, the uniform germ plasm carrying this trait (Texas [T] male-sterile cytoplasm) proved very susceptible to Southern Corn Leaf Blight. Indeed, in 1970, virtually all of the U.S. hybrid maize crop incorporated the male sterility factor, and 15 to 20 percent of the entire crop was lost. Researchers and the seed industry responded by returning to the more laborious method of detasseling by hand until new male-sterile varieties could be developed (National Research Council 1972).

Since the 1940s, international agricultural "campaigns against hunger" have been transferring maize-breeding technologies (especially those involving hybrid seed) to developing countries. Henry Wallace, mentioned previously, spearheaded the Rockefeller Foundation's agricultural research programs in Mexico and India, both of which emphasized hybrid maize. Nevertheless, in Mexico the maize agricultural sector remains mostly small-scale, semisubsistent, and traditional, and in much of Latin America, the public seed sector has been unreliable in generating and supplying improved seeds for small farmers working diverse environments; probably no more than 20 percent of Mexico's, 30 percent of Central America's, and 15 percent of Colombia's maize production has resulted from modern improved varieties (Jaffe and Rojas 1994). The Puebla Project of Mexico, which aimed to double the maize yields of small farmers by providing improved seeds, chemical packages, and a guaranteed market and credit, had only spotty participation as maize farming competed unsuccessfully with nonfarming occupations.

Agricultural research programs in British colonial Africa in the 1930s also emphasized the development of hybrid maize seed, an emphasis that was revived in Kenya, Zimbabwe, and Malawi in the 1960s and became very important in the 1980s (Eicher 1995). Zimbabwe released its first hybrid in 1949, and Kenya did the same with domestically produced hybrids in 1964. An advanced agricultural infrastructure in these countries has meant that the rate of adoption of hybrids is extremely high, and in Zimbabwe, the yields achieved by some commercial farmers approach those seen in Europe and the United States. In Ghana, the Global 2000 agricultural project, supported by the Sasakawa Africa Association, has sought to demonstrate that high yields are attainable if farmers can be assured quality seeds, affordable fertilizers, and market access. The success of hybrids in these contexts depends on timely and affordable delivery of seed and other inputs, particularly fertilizers. Tanzania, from the late 1970s through the 1980s, provided a case study of deteriorating maize production associated with erratic seed supply, elimination of fertilizer subsidies, inadequate market transportation, and insufficient improvement of open-pollinated varieties (Friis-Hansen 1994).

Under optimal conditions, hybrids yield 15 to 20 percent more than the improved open-pollinated varieties, and breeders find it easier to introduce particular traits – such as resistance to a specific disease – into inbred hybrid lines. The uniform size and maturation rate of hybrids are advantages for farmers who wish to harvest and process a standard crop as a single unit. From a commercial standpoint, hybrids also carry built-in protection against multiplication because the originator controls the parent lines, and the progeny cannot reproduce the parental type.

Conditions are rarely optimal, however, and a corresponding disadvantage is that the yields of hybrid seeds are unpredictable where soil fertility, moisture, and crop pests are less controlled. Although the introduction of disease-resistant traits may be easier with hybrids, the very uniformity of the inbred parent lines poses the risk of large-scale vulnerability, as illustrated by the case of the Southern Corn Leaf Blight. Rapid response by plant breeders and seed companies to contain the damage is less likely in countries without well-organized research and seed industries. Farmers in such countries may also face shortages of high-quality seed or other inputs, which potentially reduces the yield advantage of hybrid seed. Indeed, in years when cash is short, farmers may be unable to afford the purchase of seed or other inputs, even when available, and in any event, they lack control over the price and quality of these items – all of which means a reduction in farmer self-reliance. Analysts of public and private agricultural research systems further argue that the elevated investment in hybrids reduces the funds available for improving open-pollinated varieties and that some of the yield advantage of hybrids may result from greater research attention rather than from any intrinsic superiority.

Agricultural Research in Developing Countries

In 1943, the Rockefeller Foundation, under the leadership of Norman Borlaug, launched the first of its “campaigns against hunger” in Mexico, with the aim of using U.S. agricultural technology to feed growing populations in this and other developing countries. In 1948, the campaign was extended to Colombia, and in 1954 the Central American Maize Program was established for five countries. In 1957, the program added a maize improvement scheme for India, which became the Inter-Asian Corn Program in 1967. The Ford and Rockefeller Foundations together established the International Center for the Improvement of Maize and Wheat in Mexico in 1963, the International Center for Tropical Agriculture in Colombia in 1967, and the International Institute for Tropical Agriculture in Nigeria in 1967. These centers, with their maize improvement programs, became part of the International Agricultural Research Center Network of the Consultative Group on International Agricultural Research, which was established with coordination among the World Bank, and the Food and Agriculture Organization of the United Nations (FAO) in 1971. The International Plant Genetic Research Institute, also a part of this system, collects and preserves maize germ plasm.

Additional international maize research efforts have included the U.S. Department of Agriculture (USDA)–Kenyan Kitale maize program instituted during the 1960s; the French Institute for Tropical Agricultural Research, which works with scientists in former French colonies in Africa and the Caribbean region; and the Maize Research Institute of Yugoslavia.

The Inter-American Institute of Agricultural Sciences (Costa Rica), Centro de Agricultura Tropical para Investigación y Enseñanzas (Costa Rica), Safgrad (Sahelian countries of Africa), Saccar (southern Africa), Prociandino (Andes region), and Consasur (Southern Cone, South America) are all examples of regional institutions for maize improvement (Dowswell et al. 1996).

Cultural History

Middle and South America

In the United States, which is its largest producer, maize is business, but for indigenous Americans maize more often has been considered a divine gift, “Our Mother” (Ford 1994), “Our Blood” (Sandstrom 1991), and what human beings are made of (Asturias 1993). Archaeological evidence and ethnohistorical accounts indicate that ancient American civilizations developed intensive land- and water-management techniques to increase production of maize and thereby provision large populations of craftspeople and administrators in urban centers. Ethnohistory and ethnography depict the maize plant in indigenous thought to be analogous to a “human being,” and lexica maintain distinctive vocabularies for the whole plant, the grain, foods prepared from the grain, and the plant’s parts and life stages (seedling, leafing, flowering, green ears, ripe ears), which are likened to those of a human. Indigenous terms and usage symbolically identify the maize plant and field (both glossed in the Spanish *milpa*) with well-being and livelihood. In addition, the four principal maize kernel colors constitute the foundation of a four-cornered, four-sided cosmology, coded by color.

An inventive indigenous technique for maize preparation was “nixtamalization” (alkali processing). Soaking the grain with crushed limestone, wood ash, or seashells helped loosen the outer hull, which could then be removed by washing. This made the kernel easier to grind and to form into a nutritious food end product, such as the tortilla in Mexico and Central America or the distinctive blue *piki* bread of the U.S. Southwest. In South America, maize was also consumed as whole-grain *mote*.

Tortillas, eaten along with beans and squash seeds (the “triumvirate” of a Mesoamerican meal), constitute a nutritious and balanced diet. In Mexico and Central America, dough is alternatively wrapped in maize sheaths or banana leaves. These steamed maize-dough *tamales* sometimes include fillings of green herbs, chilli sauce, meat, beans, or sugar. Additional regional preparations include gruels (*atoles*), prepared by steeping maize in water and then sieving the liquid (a similar dish in East Africa is called *uji*); ceremonial beverages made from various maize doughs (*pozole*, which can also refer to a corn stew with whole grains, or chocolate *atole*); and special seasonal and festival foods prepared from immature

maize (including spicy-sweet *atole* and *tamales*). Green corn – a luxury food for those dependent on maize as a staple grain (because each ear consumed in the immature stage limits the mature harvest) – can be either roasted or boiled in its husk.

Andean populations also made maize beers (*chicha*) of varying potency, which involved soaking and sprouting the grain, then leavening it by chewing or salivation (Acosta 1954). Brewed maize comprised a key lubricant of Incan social life (Hastorf and Johannessen 1993). Unfortunately, by the early period of Spanish occupation, indigenous leaders were reported to be having difficulty controlling intoxication, a problem heightened when *chicha* was spiked with cheap grain alcohol – a Spanish introduction.

Other special indigenous preparations included green corn kernels boiled with green lima beans, a dish introduced to the English on the East Coast of North America. The Hopi of the North American Southwest prepared *piki*, or “paper-bread,” from a fine cornmeal batter spread on a stone slab lubricated with seed oil (from squash, sunflower, or watermelon seeds). They colored their cornbread a deep blue (or other colors) by adding extra alkalies and other pigments.

All parts of the maize plant were used by indigenous peoples. Tender inner husks, young ears, and flowers were boiled as vegetables, and fresh silks were mixed into tortilla dough. The Navajo prepared a soup and ceremonial breads from maize pollen; sugary juices were sucked out of the pith and stems; and even the bluish-black smut, *Ustilago maydis*, was prepared as a “mushroom” delicacy. Maize ear-sheaths wrapped tamales, purple sheaths colored their contents, and husks served as wrappers for tobacco. Maize vegetation was put into the food chain, first as green manure and, after Spanish livestock introductions, as animal fodder. The dried chaff of the plant was shredded for bedding material, braided into cord and basketry, and used to make dolls and other toys. Corn silks were boiled into a tea to relieve urinary problems, and the cobs served as stoppers for jugs, or as fuel. The stalks provided both a quick-burning fuel and construction material for shelters and fences. These uses continued during the Spaniards’ occupation, which added poultry and ruminants as intermediary consumers of maize in the food chain and animal manure as an element in the nitrogen cycle. Indigenous peoples generally ate maize on a daily basis, reserving wheat for festival occasions, whereas those of Spanish descent raised and consumed wheat as their daily bread.

North America

In contrast to the Spaniards, English settlers in North America adopted maize as the crop most suited to survival in their New World environment and learned from their indigenous neighbors how to plant, cultivate, prepare, and store it. Although seventeenth-century Europeans reviled maize as a food fit only for

desperate humans or swine (Gerard 1597; Brandes 1992), in North America the first colonials and later immigrants elevated the crop to the status of a staple food. For all Americans, it became a divine gift, a plentiful base for “typical” national and regional dishes, and a crop for great ritual elaboration, annual festivals, husking bees, shows, and later even a “Corn Palace” built of multicolored cobs (in Mitchell, North Dakota). In the process, maize became “corn,” originally the generic term for “grain,” shortened from the English “Indian corn,” a term that distinguished colonial exported maize (also called “Virginia wheat”) from European wheat and other grains.

Corn nourished the U.S. livestock industry, the slave economy, and westward expansion. It served as the foundation of the typical U.S. diet – high in meat and dairy products, which are converted corn – and, indeed, of the U.S. agricultural economy. North American populations of European and African ancestry historically turned maize into breads, grits, and gruels. They ate corn in the forms of mush; “spoon bread” (a mush with eggs, butter, and milk); simple breads called “hoecakes” (or “pone”); whole grains in “hominy”; and mixed with beans in “succotash.” Coarsely ground “grits” were boiled or larded into “crackling bread,” “scrapple,” “fritters,” and “hush puppies,” or were sweetened with molasses and cooked with eggs and milk into “Indian pudding.” Culinary elaborations of green corn, for which special varieties of sweet corn were bred, ranged from simple roasted (which caramelizes the sugar) or boiled “corn on the cob” with butter, to chowders and custards. Scottish and Irish immigrants fermented and distilled corn mash into corn whiskey (“white lightning” or “moonshine”) or aged and mellowed it into bourbon, a distinctively smooth American liquor named for Bourbon, Kentucky, its place of origin.

Nineteenth-century food industries added corn syrup, oil, and starch to the processed repertoire, and then corn “flakes,” the first of a series of breakfast cereals that promoters hoped would improve the healthfulness of the American diet. Popcorn, a simple indigenous food, by the mid-nineteenth century was popped in either a fatted frying pan or a wire gauze basket, and by the end of the century in a steam-driven machine, which added molasses to make Cracker Jacks, a popular new American snack first marketed at the 1893 Columbian Exposition in Chicago. By the late twentieth century, popcorn had become a gourmet food, produced from proprietary hybrid varieties, such as Orville Redenbacher’s Gourmet Popping Corn, which boasted lighter, fluffier kernels and fewer “dud” grains and could be popped in a microwave (Fussell 1992). Twentieth-century snack foods also included corn “chips” (tortillas fried in oil). Moreover, North Americans consume large quantities of corn products as food additives and ingredients, such as corn starch, high-fructose syrup, and corn oil, as well as in animal products.

Europe

Maize, introduced by Christopher Columbus into Spain from the Caribbean in 1492–3, was first mentioned as a cultivated species in Seville in 1500, around which time it spread to the rest of the Iberian peninsula. It was called *milbo* (“millet” or “grain”) by Portuguese traders, who carried it to Africa and Asia (the name survives in South African “mealies,” or cornmeal).

Spreading across Europe, maize acquired a series of binomial labels, each roughly translated as “foreign grain”: In Lorraine and in the Vosges, maize was “Roman corn”; in Tuscany, “Sicilian corn”; in Sicily, “Indian corn”; in the Pyrenees, “Spanish corn”; in Provence, “Barbary corn” or “Guinea corn”; in Turkey, “Egyptian corn”; in Egypt, “Syrian dourra” (i.e., sorghum); in England, “Turkish wheat” or “Indian corn”; and in Germany, “Welsch corn” or “Bactrian typha.” The French *blé de Turquie* (“Turkish wheat”) and a reference to a golden-and-white seed of unknown species introduced by Crusaders from Anatolia (in what turned out to be a forged Crusader document) encouraged the error that maize came from western Asia, not the Americas (Bonafous 1836). De Candolle (1884) carefully documented the sources of the misconception and also dismissed Asian or African origins of maize on the basis of its absence from other historical texts. But inexplicably, sixteenth- and seventeenth-century herbalists appear to describe and illustrate two distinct types of maize, one “definitely” from tropical America, the other of unknown origin (Finan 1950).

English sources, especially J. Gerard’s influential *Herball* (1597: Chapter 14), assessed “Turkie corne” to be unnourishing, difficult to digest, and “a more convenient food for swine than for man.” Such disparagement notwithstanding, climate and low-labor requirements for its cultivation favored maize’s dispersal. By the end of the sixteenth century, it had spread from southern Spain to the rest of the Iberian peninsula, to German and English gardens, and throughout Italy, where, by the seventeenth century, it constituted a principal element of the Tuscan diet. In both northwestern Iberia and northern Italy, climate favored maize over other cereals and gave rise to cuisines based on maize breads (*broa* and *borona*) and *polenta*. By the eighteenth century, maize had spread across the Pyrenees and into eastern France, where it became a principal peasant food and animal fodder.

A century earlier, maize had penetrated the Balkan Slavonia and Danube regions, and Serbs were reported to be producing *cucurutz* (maize) as a field crop at a time when other grains were scarce. By the mid-eighteenth century, it was a staple of the Hapsburg Empire, especially in Hungary. By the end of the eighteenth century, fields of maize were reported on the route between Istanbul and Nice, and it had likely been an earlier garden and hill crop in Bulgaria. Maize appears to have entered Romania in the beginning of the seventeenth century, where it became established

as a field crop by midcentury. T. Stoianovich (1966) traces the complex of Greek-Turkish and Romanian-Transylvanian names for maize across the region and shows how the crop was incorporated into spring planting and autumn harvest rites of local peoples. In these regions, maize, which was more highly productive, replaced millet and, especially in Romania, has been credited with furthering a demographic and agricultural-socioeconomic transition. In the nineteenth century, Hungary was a major producer, along with Romania; by the mid-1920s, the latter country was the second largest exporter of maize (after Argentina) and remained a major producer and exporter through 1939. Romania maintained its own research institute and developed its own hybrids (Ecaterina 1995).

In contrast to the potato – the other major crop from the New World – maize appears to have been introduced across Europe with little resistance or coercion. In Spain, Italy, and southern France, its high seed-to-harvest ratio, relatively low labor requirements, high disease resistance, and adaptability allowed the plant to proceed from botanical exotic to kitchen-garden vegetable to field crop, all within a hundred years (Langer 1975). Throughout Europe, maize was prepared as a gruel or porridge because its flour lacked the gluten to make good leavened bread, although it was sometimes mixed into wheat flour as an extender. Although the custom of alkali processing, or that of consuming maize with legumes, had not accompanied the plant from the New World to the Old, maize provided a healthy addition to the diet so long as consumers were able to eat it with other foods. In the best of circumstances, it became a tasty culinary staple. In the worst, undercooked moldy maize became the food of deprivation, the food of last resort for the poor, as in Spain, where for this reason, maize is despised to this day (Brandes 1992).

Curiously, maize was never accepted in the British realm, where it continued to be “an acquired taste,” a sometime famine or ration food, or a grain to feed livestock. During the great Irish famine of 1845, the British government imported maize for food relief – to keep down the prices of other foods and provide emergency rations for the poor. Maize boasted the advantages of being cheap and having no established private “free trade” with which government imports might interfere. Unfortunately, Ireland lacked the milling capacity to dry, cool, sack, and grind it, and in 1846 a scarcity of mills led to the distribution of the whole grain, which was described as irritating rather than nourishing for “half-starving people” (Woodham-Smith 1962). Maize shortly thereafter began to play an important role in British famine relief and as ordinary rations for workers in Africa.

Africa

Portuguese traders carried maize to eastern Africa in the sixteenth century, and Arab traders circulated it around the Mediterranean and North Africa. During

the seventeenth century, maize traveled from the West Indies to the Gold Coast, where, by the eighteenth century, it was used as a cheap food for provisioning slaves held in barracoons or on shipboard during the Middle Passage. By the end of the eighteenth century, maize was reported in the interior of Africa (the Lake Chad region of Nigeria), where it appears to have replaced traditional food plants in the western and central regions, especially the Congo, Benin, and western Nigeria, although cassava – because it was less vulnerable to drought and locusts – later replaced maize in the southern parts of Congo (Miracle 1966) and Tanzania (Fleuret and Fleuret 1980).

By the end of the nineteenth century, maize had become established as a major African crop. People accustomed to eating it as the regular fare in mines or work camps, or as emergency rations, now demanded it when conditions returned to normal or when they returned home. Consumption also increased following famine years because people were able to sow earlier-maturing varieties, and even where sorghum remained the principal staple, maize became a significant seasonal food that was consumed at the end of the “hungry season” before other cereals ripened. The British also promoted African maize as a cash crop for their home starch industry.

Today in Africa, ecology, government agricultural and marketing policies, and the cost of maize relative to other staple or nonstaple crops are the factors influencing the proportion of maize in household production and consumption (Anthony 1988). Major shifts toward maize diets occurred in the latter part of the twentieth century, when improved varieties and extension programs, as well as higher standards of living, meant that people could enjoy a more refined staple food – with less fiber – without feeling hungry. Researchers in postcolonial times have developed hybrids adapted to African conditions, but these have met with mixed reactions. In Zimbabwe, where small farmers are well organized and can demand seed and access to markets, most of them plant improved hybrids (Bratton 1986; Eicher 1995). By contrast, in Zambia, smallholders continue to grow traditional varieties for a number of reasons, including the high cost of hybrid seed, shortages of seed and input supplies, inadequate storage facilities, and a culinary preference for varieties that are flintier. The latter are preferred because they have a higher extraction rate when mortar-pounded (superior “mortar yield”), they produce superior porridge, and they are more resistant to weevils. However, even where introduction of improved disease-resistant varieties has been successful, as in northern Nigeria, the gains will be sustainable only if soils do not degrade, the price of fertilizer remains affordable, markets are accessible, and research and extension services can keep ahead of evolving pests (Smith et al. 1994).

African cuisines favor white maize, which is prepared as a paste or mush and usually eaten as warm

chunks dipped in stews or sauces of meat, fish, insects, or vegetables. In eastern and southern Africa, maize is first pounded or ground before being boiled into a thick porridge. But in western Africa, *kenkey* is prepared from kernels that are first soaked and dehulled before being ground, fermented, and heated. *Ogi* is a paste prepared by soaking the kernels, followed by light pounding to remove the pericarp and a second soaking to produce a bit of fermentation. The bran is strained away, and the resulting mass cooked into a paste, mixed with the dough of other starchy staples, and baked into an unleavened bread or cooked in oil. Maize gruels can be soured or sweetened, fermented into a light or a full beer, or distilled. The kernels are also boiled and eaten whole, sometimes with beans, or beaten into a consistency like that of boiled rice. Alternatively, the grains can be parched before boiling, or cooked until they burst. Immature ears are boiled or roasted, and the juice from immature kernels flavored, cooked, and allowed to jell.

Asia

Portuguese introductions of maize to Asia likely occurred in the early 1500s, after which the grain was carried along the western coast of India and into northwestern Pakistan along the Silk Route. By the mid-1500s, maize had reached Honan and Southeast Asia, and by the mid-1600s it was established in Indonesia, the Philippines, and Thailand. In southern and southwestern China during the 1700s, raising maize permitted farming to expand into higher elevations that were unsuitable for rice cultivation, and along with white and sweet potatoes, the new crop contributed to population growth and a consequent growing misery (Anderson 1988). From there, maize spread to northern China, Korea, and Japan.

In the 1990s, maize was a staple food in selected regions of China, India, Indonesia, and the Philippines. Among grains in China, it ranked third in importance (after rice and wheat, but before sorghum, which it has replaced in warmer, wetter areas and in drier areas when new hybrids became available). Maize is consumed as steamed or baked cakes, as mush, in noodles mixed with other grains, and as cracked grain mixed with rice. It is the principal grain in the lower mountains of western and southern China and in much of the central North, and it increasingly serves as a food for the poor. Immature maize is eaten as a vegetable, and baby corn is an important specialty appreciated for its crunchy texture.

In Indonesia, maize is the staple food of some 18 million people (Timmer 1987). Farmers have responded favorably to new technologies and government incentives, such as quick-yielding varieties, subsidized fertilizers, and mechanical tilling and shelling devices. They demand improved seed and subsidized chemicals and carefully match seed varieties to local seasonal conditions in environments that (in places)

allow for triple cropping (rice-maize-maize, rice-maize-soy, or rice-maize-cassava sequences). In the 1980s, breeders reportedly could not keep up with the demand for improved white varieties, which cover 35 percent of the area sown in maize and are preferred for human consumption. Humans still consume 75 percent of Indonesia's maize directly, and it is particularly important as a staple in the preharvest "hungry season" before the main rice harvest. Rice remains the preferred staple; the proportion of maize in the diet fluctuates relative to the price of maize versus rice and consumer income.

Summary of Culinary History

Kernels of the earliest forms of maize were probably parched on hot stones or in hot ash or sand. Small hard seeds, in which starch was tightly packed, also lent themselves to popping. Both Mexican and Peruvian indigenous populations grew selected popcorn varieties, which, among the Aztecs, were burst like flowers or hailstones for their water god. Parched maize, sometimes mixed with other seeds, was ground into *pinole*, a favorite lightweight ration for travelers that could be eaten dry or hydrated with water.

Maize grains more commonly were wet-ground - boiled and then ground with a stone quern or pounded with wooden implements (the North American indigenous procedure outside of the Southwest). After soaking the kernels in alkaline and washing to remove the hulls, native peoples either consumed the grains whole (as hominy) or wet-ground them into a dough used to form tortillas (flat cakes), *arepas* (thick cakes), or *tamales* (leaf-wrapped dough with a filling). The arduous process of grinding, which could require up to half of a woman's workday, was later taken over by water- or engine-powered mills; the time-consuming process of shaping tortillas by hand was facilitated by wooden or metal tortilla "presses"; and very recently, the entire process of tortilla manufacture has been mechanized. In 1995, the people of Mexico consumed 10 million tons of tortillas, each ton using 10,000 liters of water to soak and wash the kernels, water that, when dumped, created rivers of calcium hydroxide. An interdisciplinary team has formed a 1995 "Tortilla Project" to create a water-saving, energy-efficient machine that will turn out a superior nutritional product with no pollutants.

Dry-grinding, characteristic of nonindigenous processing, produces whole maize "meal," "grits," or "flour," which can be "decorticated" (bran removed) or "degerminated" (most bran and germ removed), a separation process also called "bolting," to produce a more refined flour and an end product that stores longer. Hominy is the endosperm product left over after the pericarp is removed and the germ loosened; "pearl" or "polished" hominy has the aleurone layer removed as well. Although separating the bran and germ decreases the vitamin and mineral value, it

makes the oil and residual "germ cake," pericarp, and hulls more easily available for livestock feed.

Simple, boiled maize-meal porridges, which combine whole or degermed meal with water, are the most common forms of maize dishes. In the United States, it is cornmeal mush; in Italy, *polenta*; in Romania, *mamaliga*; and in Africa, *nshima*, *ugali*, or *foo foo*. In Italy, *polenta* often includes grated cheese and extra fat, and in Yugoslavia, the corn mush contains eggs and dairy products. Maize meal, when mixed with water or milk, can be shaped into simple unleavened flat breads or cakes and baked over an open fire, or in an oven. These were called *boecakes* in the early United States.

In Asia, maize is "riced"; the cracked kernels are boiled and consumed like the preferred staple of the region. Indeed, improved maize varieties must meet the processing criteria of cracking and cooking to resemble rice. Maize starch, especially in central Java, is processed into flour for porridge, noodles, and snack food. Green maize is also consumed.

Industrial Processing

Industrial processing utilizes either wet or dry milling. Wet milling steeps kernels in water, separates germ from kernel, and then separates germ into oil and meal portions. Each 100 kilograms (kg) of maize yields 2 to 3 kg of oil. Corn oil is popular for its ability to withstand heat, its high level of polyunsaturates, and its flavorlessness. The meal portions of the kernel become starch, gluten, and bran.

The dried starch portion is used in the food, textile, and paper industries. Starch, processed into glucose syrup or powder and high-fructose or dextrose products, sweetens three-fourths of the processed foods in the United States, especially confections and baked goods. In 1991, high-fructose corn syrup, manufactured by an enzyme-isomerization process that was first introduced in the 1960s, accounted for more than half of the U.S. (nondiet) sweetener market (National Corn Growers Association 1992).

Dry milling processes about 2 percent of the U.S. maize crop into animal feeds, beers, breakfast cereals, and other food and industrial products. In a tempering/degerming process the majority of the pericarp and germ are removed, and the remaining bulk of the endosperm is dried and flaked into products such as breakfast cereal. Whole (white) grains are ground into hominy grits and meal. These products, because they still contain the oily germ, have superior flavor but shorter shelf life. Industrialized alkali processing produces a dough that is turned into tortillas, chips, and other "Mexican" snacks.

Special maize varieties are also being bred for "designer" industrial starches. One, high in amylose starch, is used to create edible wrappers for pharmaceuticals, feeds, and foods. Another "super slurper" absorbs up to 2,000 times its weight in moisture and

is employed in disposable diapers and bedpads. Still another is being developed into less-polluting road "salt," and other corn starches are being tailored into biodegradable plastic bags, cups, and plates. All told, industrial maize preparations place thousands of different maize-derived items on modern supermarket shelves, including flours and meals for breads and puddings, starch as a thickener, maize ("Karo") syrups or honeys as sweeteners, high-fructose and dextrose syrups as sweetening ingredients in beverages and baked goods, and processed cereals as breakfast or snack foods. Maize-based cooking oils, chips, beers, and whiskeys complete the spectrum. In fact, almost all processed foods (because they contain additives of starch or fat) contain some maize, as do almost all animal products, which are converted maize (Fussell 1992).

Animal Feed Products

Maize is the preferred feedgrain for animals because it is so rich in fat and calories; its high-starch/low-fiber content helps poultry, swine, cattle, and dairy animals convert its dry matter more efficiently than with other grains; it is also lower in cost. Feeds are formulated from whole, cracked, or steam-flaked grains and optimally supplemented with amino acids, vitamins, and minerals to meet the special nutritional requirements of particular domesticated animals. In industrial processing, by-products remaining after the oil and starch have been extracted – maize gluten, bran, germ meal, and condensed fermented steepwater (from soaking the grain), which is a medium for single-cell protein – also go into animal feed.

Silage uses the entire maize plant – which is cut, chopped, and allowed to ferment – to nourish dairy and beef cattle and, increasingly, swine. In developing countries, fresh or dried vegetation and substandard grains are household commodities used to produce animal products. When the entire maize plant (and, in traditional fields, its associated weeds) serves as a feedstuff, it surpasses all other plants in average yield and digestible nutrients per hectare (Dowswell et al. 1996).

Nutrition

Maize provides 70 percent or more of food energy calories in parts of Mexico, Central America, Africa, and Romania. In these regions, adult workers consume some 400 grams of maize daily, a diet marginally sufficient in calories, protein, vitamins, and minerals, depending on how the maize is processed and the supplementary foods with which it is combined. Maize is a better source of energy than other cereal grains because of its higher fat content. Ground maize meal has 3,578 calories per kg, mostly carbohydrate, with about 4.5 percent "good" fat (fat-rich varieties are double this figure), and is high in essential linoleic and oleic fatty acids. It contains about 10 percent pro-

tein, roughly half of which is zein that is low in the amino acids lysine and tryptophan. The protein quality is enhanced in traditional Latin American maize diets by alkali processing and consumption with legumes that are high in lysine. Potatoes, if eaten in sufficient quantity, also yield a considerable amount of lysine and consequently often complement maize in highland South America and parts of Europe. Of course, incorporating animal proteins improves the nutritional quality of any diet with grain or tubers as the staple.

Maize is also naturally low in calcium and niacin, but calcium, niacin, and tryptophan content are all enhanced by traditional alkali processing (in which the kernels are cooked and soaked in a calcium hydroxide – lime or ash – solution), which adds calcium and increases the available tryptophan and niacin in the kernels or dough. White maize, usually the favored type for human food, is also low in vitamin A, although this nutrient is higher in properly stored yellow maize. Moreover, in its traditional heartland, maize is combined with chilli peppers, other vegetables, and various kinds of tomatoes and spices, all of which enhance the amount of vitamin A delivered to the consumer, along with other vitamins and minerals. In Africa and Asia, additional vitamins and minerals are added to maize diets when wild or cultivated greens, other vegetables, peanuts, and small bits of animal protein are combined in a sauce. Potash, burned from salt grasses, also enhances the otherwise poor mineral content of the diet (FAO 1953).

Diseases Associated with Maize

Pellagra and protein-deficiency disease (kwashiorkor) are historically associated with maize diets. In addition, as recently as the 1950s, rickets, scurvy, and signs of vitamin A deficiency have all been reported among populations consuming maize diets in Central Europe and eastern and southern Africa. Such deficiency diseases disappear with dietary diversification, expansion of food markets, and technological efforts to improve the micronutrient quality of maize diets.

Pellagra

Pellagra, now understood to be a disease caused in large part by niacin deficiency, was first observed in eighteenth-century Europe among the very poor of Spain, then Italy, France, Romania, Austria, southern Russia, the Ottoman Empire, and outside of Europe in Egypt, South Africa, and the southern United States. It was associated with extreme poverty and usually seen among land-poor peasants, whose diet centered much too heavily on maize. The main symptoms were described as the "three Ds" (diarrhea, dermatitis, and dementia), and four stages were recognized, from malaise, to digestive and skin disorders, to neurological and mental symptoms, and finally, wasting, dementia, and death (Roe 1973).

Although maize was adopted as a garden crop and within 100 years after its appearance was a field crop over much of the European continent, the disease manifested itself only when economic conditions had deteriorated to the point that pellagra victims ("pellagrins") could afford to eat only poorly cooked, often rotten maize. In Spain, this occurred in the 1730s; up to 20 percent of the population may still have been afflicted in the early twentieth century. In Italy, peasants also may have been suffering from the "red disease" as early as the 1730s. Despite efforts to protect the purity of the maize supply and improve diets through public granaries, bakeries, and soup kitchens, the disease persisted until the 1930s, when changes in diet were brought about by improved standards of living and the demise of the tenant-farmer system. In France, where maize had been sown since the sixteenth century and in some areas had expanded into a field crop by the late seventeenth, maize was not widely grown as a staple until the late eighteenth and early nineteenth centuries, when it became the main crop of the southern and eastern regions of the country and was accompanied by pellagra among destitute peasants. The physician Theophile Roussel recommended that the disease be prevented by changing the diet and agriculture so that there was less emphasis on maize. The government responded with legislation encouraging alternative crop and livestock production along with consumption of wheat, and by the early twentieth century, pellagra had largely been eliminated.

In the late nineteenth century, pellagra was reported by a British physician, Fleming Sandwith, in Egypt and South Africa. The disease was also present in the southern United States, although it did not reach epidemic proportions until the first decade of the twentieth century. In epidemiological studies begun in 1914, Joseph Goldberger, a physician working for the Public Health Service, determined that the disease was not contagious but *was* dietary. Furthermore, it was associated not so much with the consumption of maize as with the economic inability to obtain and consume other protective foods along with maize. For prevention and cure, Goldberger prescribed milk, lean meat, powdered yeast, and egg yolks. At the household level, he recommended more diversified farming, including milk cows and more and better gardens.

Goldberger traced the correlation between epidemic pellagra and economic downturns and demonstrated how underlying socioeconomic conditions restricted diets and caused dietary deficiencies among tenant farmers, who ordinarily ate mostly maize and maize products. The number of cases declined in the worst depression years (1932–4) because, when there was no market for cotton, farmers produced diversified food crops and garden vegetables for home consumption. Goldberger also demonstrated that pellagra mimicked "blacktongue" in

dogs and used them as experimental animals to find what foods might prevent pellagra. He conceptualized the "pellagra-preventive" factor to be a water-soluble vitamin but could not identify it (Terris 1964). It was not until 1937 that C. A. Elvehjem and his colleagues demonstrated that nicotinic acid cured blacktongue in dogs, a finding carried over to demonstrate that nicotinic acid prevented pellagra in humans. Lest the public confuse nicotinic acid with nicotine, the Food and Drug Administration adopted the name "niacin" for their vitamin fortification program (Roe 1973: 127), which was designed to eliminate nutrition-deficiency diseases, and in southern states tended to include cornmeal and grits as well as wheat flours. Diversification and improvement of diet associated with World War II production, employment, and high-quality food rations mostly spelled an end to pellagra in the United States.

Since the 1940s, maize diets and pellagra have also been associated with imbalanced protein intake and selected amino acid deficiency. G. A. Goldsmith (1958) demonstrated that dietary tryptophan converts to nicotinic acid in humans at a ratio of 1:45, and anthropologists working with nutritional chemists were able to demonstrate that alkali processing of maize in traditional indigenous diets made more niacin and tryptophan available (Katz, Hediger, and Valleroy 1974). Traditional processing and food combinations also make more isoleucine available relative to leucine, and it has been suggested that excess leucine is another antinutritional factor in maize. Although pellagra has been eliminated in industrialized countries, it remains a plague among poor, maize-eating agriculturalists in southern Africa, where it was reported throughout the 1960s in South Africa, Lesotho, and Tanzania, and in Egypt and India among people who lack access to wheat.

Protein Deficiency

Another nutritional deficiency disease historically associated with diets high in maize is kwashiorkor, conventionally classified as a protein-deficiency disease and associated especially with weanlings and hungry-season diets in Africa (Williams 1933, 1935). Since the 1960s, international maize-breeding programs have sought to overcome lysine deficiency directly, thus giving maize a much better-quality protein. Maize breeders at Purdue University in Indiana, who were screening maize for amino acid contents, isolated the mutant "Opaque-2" gene and developed a variety that had the same protein content as conventional maizes but more lysine and tryptophan.

Although this variety possessed a more favorable amino acid profile, its yields were lower, its ears smaller, its chalky kernels dried more slowly, and it carried unfavorable color (yellow), texture, and taste characteristics. Its softer, more nutritious, and moister starch was more easily attacked by insects and fungi, and its adhesive properties did not make a good

tortilla. Mexican researchers at CIMMYT in the 1970s and 1980s eliminated these deficiencies and in the mid-1980s introduced Quality Protein Maize (QPM) with favorable consumer characteristics. The remaining step was to interbreed this superior type with locally adapted varieties. But by the 1980s, nutritionists were questioning the importance of protein or selective amino-acid deficiencies as high-priority problems and focusing instead on improving access to calories. QPM became a technological solution for a nutritional deficiency no longer of interest, and CIMMYT was forced to end its QPM program in 1991. However, national programs in South Africa, Ghana, Brazil, and China are using QPM to develop maize-based weaning foods and healthier snacks, as well as a superior animal feed (Ad Hoc Panel 1988).

Additional Strategies for Nutritional Improvement

Strategies for improving maize diets focus on new varieties with higher protein quality and essential vitamin contents, better storage, wiser milling and processing, fortification, and dietary diversification. Conventional breeding and genetic engineering have enhanced essential amino acid profiles, especially lysine and methionine contents, although end products so far are principally superior feeds for poultry and pigs. Maize transformation by means of electroporation and regeneration of protoplasts was achieved in 1988, and subsequently by *Agrobacterium* (Rhodes et al. 1988). The first commercial varieties, with added traits of herbicide resistance and superior protein quality, were released in 1996. To improve protein content, maize meals are also fortified with soybean protein meal or dried food yeast (*Tortula utilis*). Nutritional enhancement through breeding or blending are alternatives to diversifying human (or animal) diets with legumes or animal foods.

Improperly stored maize, with moisture contents above 15 percent, also favor the growth of fungi, the most dangerous being *Aspergillus flavus*, which produces aflatoxin, a mycotoxin that causes illness in humans and animals. Efforts are being taken to eliminate such storage risks and losses.

Future Prospects

Maize has been expanding in geographical and cultural scope to the point where the world now harvests more than 500 million tons on all continents, and the crop is being increasingly directed into a number of nonfood uses, especially in industrialized countries. The supply of maize should continue to increase in response to a growing demand for animal products (which rely on maize feed), for food ingredients industrially processed from maize (such as good-quality cooking oil), and for convenience foods and snack foods (Brenner 1991). The biologi-

cal characteristics of maize that have always favored its expansion support the accuracy of such a prediction: Its adaptability, high yields, high extraction rate, and high energy value deliver higher caloric yields per unit area than wheat or rice, and its high starch and low fiber content give the highest conversion of dry matter to animal product. Technology, especially biotechnology, will influence overall yields as well as nutritive value and processing characteristics. Genetic engineering has already allowed seed companies to market higher protein-quality maize designed to meet the specific nutritional needs of poultry and livestock. Other varieties have been designed to tolerate certain chemicals and permit higher maize yields in reduced-pest environments. The introduction of a male sterility trait, developed by Plant Genetic Systems (Belgium) in collaboration with University of California researchers, is expected to reduce the costs of manual or mechanical detasseling, estimated to be \$150 to \$200 million annually in the United States and \$40 million in Europe (Bijman 1994).

Yet, the favorable agricultural, nutritional, and economic history of maize notwithstanding, the grain presents problems. As we have seen, maize diets have been associated with poverty and illness, especially the niacin-deficiency scourge, pellagra, and childhood (weanling) malnutrition. Moreover, the highly productive inbred hybrids, such as those that contain the trait for cytoplasmic male sterility, have created new genetic and production vulnerabilities (National Research Council 1972). Hybrid seeds also may increase the economic vulnerability of small-scale semisubsistence farmers who cannot afford to take advantage of new agricultural technologies and, consequently, find themselves further disadvantaged in national and international markets. Finally, and paradoxically, maize (like the potato) has been associated with increasing hunger and suffering in Africa (Cohen and Atieno Odhiambo 1989) and Latin America (Asturias 1993).

Ellen Messer

Many thanks to Mary Eubanks, who contributed a substantial text on which the author based the section "Biological Evolution."

Bibliography

- Acosta, J. de. 1954. *Historia natural y moral de las Indias*. Madrid.
- Ad Hoc Panel of the Advisory Committee on Technology Innovation, Board on Science and Technology for International Development, National Research Council. 1988. *Quality-protein maize*. Washington, D.C.
- Allen, W. 1965. *The African busbandman*. Edinburgh.
- Anderson, E. N. 1988. *The food of China*. New Haven, Conn.

- Anghiera, P. Martire d' P. 1912. *De Orbe Novo, the eight decades of Peter Martyr d'Anghera*, trans. F. A. MacNutt. New York.
- Anthony, C. 1988. *Mechanization and maize. Agriculture and the politics of technology transfer in East Africa*. New York.
- Asturias, M. A. 1993. *Men of maize*, trans. G. Martin. Pittsburgh, Pa.
- Austin, J., and G. Esteva, eds. 1987. *Food policy in Mexico*. Ithaca, N.Y.
- Barkin, D. 1987. SAM and seeds. In *Food policy in Mexico*, ed. J. Austin and G. Esteva, 111-32. Ithaca, N.Y.
- Barkin, D., R. L. Batt, and B. R. DeWalt. 1990. *Food crops and feed crops: The global substitution of grains in production*. Boulder, Colo.
- Beadle, G. W. 1939. Teosinte and the origin of maize. *Journal of Heredity* 30: 245-7.
- Bellon, M. 1991. The ethnecology of maize variety management: A case study from Mexico. *Human Ecology* 19: 389-418.
- Bijman, J. 1994. Plant genetic systems. *Biotechnology and Development Monitor* 19: 19-21.
- Bonafous, M. 1836. *Histoire naturelle, agricole et économique du maïs*. Paris.
- Bonavia, D., and A. Grobman. 1989. Andean maize: Its origins and domestication. In *Foraging and farming: The evolution of plant exploitation*, ed. D. R. Harris and G. C. Hillman, 456-70. London.
- Brandes, S. 1992. Maize as a culinary mystery. *Ethnology* 31: 331-6.
- Bratton, M. 1986. Farmer organizations and food production in Zimbabwe. *World Development* 14: 367-84.
- Brenner, C. 1991. *Biotechnology for developing countries: The case of maize*. Paris.
- Brush, S., M. Bellon, and E. Schmidt. 1988. Agricultural diversity and maize development in Mexico. *Human Ecology* 16: 307-28.
- Candolle, A. de. 1884. *Origin of cultivated plants*. London.
- CIMMYT (International Center for the Improvement of Maize and Wheat). 1981. *World maize facts and trends. Report 1*. Mexico City.
1984. *World maize facts and trends. Report 2. An analysis of changes in Third World food and feed uses of maize*. Mexico City.
1987. *World maize facts and trends. Report 3. An analysis of the economics of commercial maize seed production in developing countries*. Mexico City.
1990. *World maize facts and trends. Realizing the potential of maize in sub-Saharan Africa*. Mexico City.
1992. *1991/1992 World maize facts and trends*. Mexico City.
- CIMMYT. 1994. *1993/1994 World maize facts and trends*. Mexico City.
- Cobo, B. 1890-3. *Historia del nuevo mundo*. Seville.
- Coe, S. 1994. *America's first cuisine*. Austin, Tex.
- Cohen, D. W., and E. S. Atieno Odhiambo. 1989. The hunger of Obalo. In *Siaya: The historical anthropology of an African landscape*, 61-84. London and Athens, Ohio.
- Cohen, J. 1995. Project refines an ancient staple with modern science. *Science* 267: 824-5.
- De Janvry, A., E. Sadoulet, and G. Gordillo de Anda. 1995. NAFTA and Mexico's maize producers. *World Development* 23: 1349-62.
- del Paso y Troncoso, F. 1905. *Papeles de Nueva Espana, Segunda Serie, geografia y estadística*. Madrid.
- De Walt, K. 1990. Shifts from maize to sorghum production: Nutrition effects in four Mexican communities. *Food Policy* 15: 395-407.
- Dowswell, C., R. L. Paliwal, and R. P. Cantrell. 1996. *Maize in the Third World*. Boulder, Colo.
- Ecaterina, P. 1995. Corn and the development of agricultural science in Romania. *Agricultural History* 69: 54-78.
- Eicher, C. 1995. Zimbabwe's maize-based Green Revolution: Preconditions for replication. *World Development* 23: 805-18.
- Eubanks, M. 1995. A cross between two maize relatives: *Tripsacum dactyloides* and *Zea diploperennis* (Poaceae). *Economic Botany* 49: 172-82.
- Fernandez de Oviedo y Valdes, G. 1526, 1530. *Historia natural y general de las Indias, islas y tierra firme del mar oceano*. Seville.
- Finan, J. 1950. *Maize in the great herbals*. Waltham, Mass.
- Fleuret, P., and A. Fleuret. 1980. Nutritional implications of staple food crop successions in Usambara, Tanzania. *Human Ecology* 8: 311-27.
- FAO (Food and Agricultural Organization of the United Nations). 1953. *Maize and maize diets*. Rome.
1993. *Average annual maize area, yield, and production, 1990-1992*. Agrostat/PC Files, Rome.
- Ford, R. 1994. Corn is our mother. In *Corn and culture in the prehistoric New World*, ed. S. Johannessen and C. Hastorf, 513-26. Boulder, Colo.
- Freeling, M., and V. Walbot, eds. 1993. *The maize handbook*. New York.
- Friis-Hansen, E. 1994. Hybrid maize production and food security in Tanzania. *Biotechnology and Development Monitor* 19: 12-13.
- Fussell, B. 1992. *The story of corn*. New York.
- Galinat, W. C. 1992. Maize: Gift from America's first peoples. In *Chilies to chocolate. Food the Americas gave the world*, ed. N. Foster and L. S. Cordell, 47-60. Tucson, Ariz.
- Gallagher, J. P. 1989. Agricultural intensification and ridged-field cultivation in the prehistoric upper Midwest of North America. In *Foraging and farming: The evolution of plant exploitation*, ed. D. R. Harris and G. C. Hillman, 572-84. London.
- Gerard, J. 1597. *The herball. Generall historie of plants*. London.
- Goldsmith, G. A. 1958. Niacin-tryptophan relationships in man and niacin requirement. *American Journal of Clinical Nutrition* 6: 479-86.
- Goodman, M. 1988. U.S. maize germplasm: Origins, limitations, and alternatives. In *Recent advances in the conservation and utilization of genetic resources*. Proceedings of the Global Maize Germplasm Workshop. Mexico.
- Harshberger, J. W. 1896. Fertile crosses of teosinte and maize. *Garden and Forest* 9: 522-3.
- Hastorf, C. A. 1994. Cultural meanings (Introduction to Part 4). In *Corn and culture in the prehistoric New World*, ed. S. Johannessen and C. A. Hastorf, 395-8. Boulder, Colo.
- Hastorf, C., and S. Johannessen. 1993. Pre-Hispanic political change - the role of maize in the central Andes of Peru. *American Anthropologist* 95: 115-38.
- Hernandez, F. 1615. *Rerum medicarum Novae Hispaniae, Thesaurus*. Mexico.
- Hernandez X., E. 1971. *Exploracion etnobotanica y su metodologia*. Mexico, D.F.
- Hewitt de Alcantara, C. 1976. *Modernizing Mexican agriculture: Socioeconomic implications of technological change, 1940-1970*. Geneva.
1992. *Economic restructuring and rural subsistence in Mexico. Maize and the crisis of the 1980s*. New York.

- Hobhouse, W. 1986. *Seeds of change: Five plants which changed the world*. New York.
- Iltis, H. 1983. From teosinte to maize: The catastrophic sexual mutation. *Science* 222: 886-93.
- Jaffe, W., and M. Rojas. 1994. Maize hybrids in Latin America: Issues and options. *Biotechnology and Development Monitor* 19: 6-8.
- Janossy, A. 1970. Value of Hungarian local maize varieties as basic material for breeding. In *Some methodological achievements of the Hungarian hybrid maize breeding*, ed. I. Kovács, 17-22. Budapest.
- Jennings, B. 1988. *Foundations of international agricultural research: Science and politics in Mexican agriculture*. Boulder, Colo.
- Jimenez de la Espada, M., ed. 1881-97. *Relaciones geograficas de las Indias*. 4 vols. Madrid.
- Johannessen, S., and C. A. Hastorf, eds. 1994. *Corn and culture in the prehistoric New World*. Boulder, Colo.
- Kato, Y. T. A. 1984. Chromosome morphology and the origin of maize and its races. *Evolutionary Biology* 17: 219-53.
- Katz, S. H., M. L. Hediger, and L. A. Valleroy. 1974. Traditional maize processing technologies in the New World. *Science* 184: 765-73.
- Kempton, J. H., and W. Popenoe. 1937. *Teosinte in Guatemala*. Contributions to American Anthropology No. 23. Washington, D.C.
- Kumar, S., and C. Siandwazi. 1994. Maize in Zambia: Effects of technological change on food consumption and nutrition. In *Agricultural commercialization, economic development, and nutrition*, ed. J. von Braun and E. Kennedy, 295-308. Baltimore, Md.
- Langer, W. 1975. American foods and Europe's population growth 1750-1850. *Journal of Social History* 8: 51-66.
- Long, J., ed. 1996. *Conquista y comida*. Mexico City.
- Mangelsdorf, P. 1974. *Corn: Its origin, evolution, and improvement*. Cambridge, Mass.
- Miracle, M. 1966. *Maize in tropical Africa*. Madison, Wis.
- Montanez, A., and A. Warman. 1985. *Los productores de maize en Mexico. Restricciones y alternativos*. Mexico City.
- National Corn Growers Association. 1992. *Annual report*. St. Louis, Mo.
- National Research Council (U.S.) Committee on Genetic Vulnerability of Major Crops, National Academy of Sciences. 1972. *Genetic vulnerability of major crops*. Washington, D.C.
- Nutall, Z. 1930. Documentary evidence concerning wild maize in Mexico. *Journal of Heredity* 21: 217-20.
- Redclift, M. 1983. Production programs for small farmers: Plan Puebla as myth and reality. *Economic Development and Cultural Change* 31: 551-70.
- Rhodes, C. A., D. Pierce, I. Mettler, et al. 1988. Genetically transformed maize plants from protoplasts. *Science* 240: 204-7.
- Richards, A. 1939. *Land, labour, and diet in northern Rhodesia: An economic study of the Bemba tribe*. London.
- Roe, D. 1973. *A plague of corn: The social history of pellagra*. Ithaca, N.Y.
- Root, W., and R. de Rochemont. 1976. *Eating in America: A history*. New York.
- Sahagun, B. de. 1950-82. *General history of the things of New Spain*. Florentine Codex. Santa Fe, N. Mex.
- Sandstrom, A. R. 1991. *Corn is our blood: Culture and ethnic identity in a contemporary Aztec Indian village*. Norman, Okla.
- Sanoja, M. 1989. From foraging to food production in north-eastern Venezuela and the Caribbean. In *Foraging and farming: The evolution of plant exploitation*, ed. D. R. Harris and G. C. Hillman, 523-37. London.
- Smale, M. 1995. "Maize is life": Malawi's delayed Green Revolution. *World Development* 23: 819-31.
- Smith, J., A. D. Barau, A. Goldman, and J. H. Mareck. 1994. The role of technology in agricultural intensification: The evolution of maize production in northern Guinea savannah of Nigeria. *Economic Development and Cultural Change* 42: 537-54.
- Sprague, G. F., and J. W. Dudley, eds. 1988. *Corn and corn improvement*. Madison, Wis.
- Stoianovich, T. 1966. Le mais dans les Balkans. *Annales* 21: 1026-40.
- Taba, S., ed. 1996. *Maize genetic resources*. Mexico.
- Terris, M., ed. 1964. *Goldberger on pellagra*. Baton Rouge, La.
- Timmer, C. P., ed. 1987. *The corn economy of Indonesia*. Ithaca, N.Y.
- Vavilov, N. I. 1931. Mexico and Central America as the principal centre of origin of cultivated plants in the New World. *Bulletin of Applied Botany, Genetics, and Plant Breeding* 26: 135-99.
- Walden, David B., ed. 1978. *Maize breeding and genetics*. New York.
- Wallace, H. A., and E. N. Bressman. [1923] 1949. *Corn and corn growing*. Fifth revised edition. New York.
- Wallace, H. A., and W. L. Brown. [1956] 1988. *Corn and its early fathers*. Revised edition. Ames, Iowa.
- Watson, P. J. 1989. Early plant cultivation in the eastern woodlands of North America. In *Foraging and farming: The evolution of plant exploitation*, ed. D. R. Harris and G. C. Hillman, 555-71. London.
- Weatherwax, P. 1954. *Indian corn in old America*. New York.
- Wellhausen, E. J., L. M. Roberts, E. Hernandez X., and P. Mangelsdorf. 1952. *Races of maize in Mexico*. Cambridge, Mass.
- Wilkes, G. 1989. Maize: Domestication, racial evolution, and spread. In *Foraging and farming: The evolution of plant exploitation*, ed. D. R. Harris and G. C. Hillman, 440-55. London.
- Wilkes, G., and M. M. Goodman. 1996. Mystery and missing links: The origin of maize. In *Maize genetic resources*, ed. S. Taba, 106. Mexico, D.F.
- Woodham-Smith, C. 1962. *The great hunger: Ireland 1845-9*. New York.

II.A.5 Millets

The caryopses of grasses have been harvested as human food since long before the advent of agriculture. Numerous species are still regularly harvested in Africa and Asia during times of scarcity. Among the many hundreds of species harvested as wild cereals, 33 species belonging to 20 genera were domesticated. Their cultivated cereals are dependent on humans for survival because they have lost the ability of natural seed dispersal and have become adapted to cultivated fields.

Cereals are grown on an estimated 730 million hectares and produce about 1,800 million metric tons of grain annually. Wheat, maize, and rice account for at least 80 percent of the annual world cereal production. Barley, sorghum, oats, rye, and pearl millet represent about 19 percent of cereal grains produced, and the remaining 1 percent of production comes from the other 19 grass species that are still grown as human food. These species are minor in terms of total world cereal production, but some are important components of agriculture in Africa and Asia (de Wet 1989).

Cereals that do not belong to the wheat (*Triticum*), barley (*Hordeum*), oats (*Avena*), maize (*Zea*), or rice (*Oryza*) genera are commonly referred to as millets (de Wet 1989).

American Millets

The first cultivated cereal in the Americas appears to have been a species of *Setaria* (Callen 1965, 1967). Archaeological records indicate that this millet was an important source of food in the Valley of Mexico and in northeastern Mexico before the domestication of maize. E. O. Callen (1967) demonstrated a steady increase in size of caryopses of this millet over 1,500 years of use as a cereal. The species, however, never lost the ability of natural seed dispersal. It was displaced by maize as a cereal during the fifth millennium B.C., but later enjoyed a temporary resurgence in importance when it was probably harvested from weed populations that invaded maize fields.

The archaeological record indicates that another native cereal was cultivated in the southeastern United States before the introduction of maize about 3,000 years ago (Wills 1988). Maygrass (*Phalaris caroliniana* Walt.) was a common component of early agricultural settlements of the region (Chomko and Crawford 1978). It has been proposed that this species was planted by the inhabitants of these early settlements, as they were located well outside the natural range of maygrass (Cowan 1978). Morphological evidence of its domestication, however, is absent.

Two native grass species besides maize, mango (*Bromus mango* Desv.), and sauwi (*Panicum sonorum* Beal) were fully domesticated in the Americas by early farming communities. Mango is the only cereal known to have become extinct in historical times (Cruz 1972). Its cultivation was confined to central Chile. In 1782, it was recorded that the Aracanian Indians of that region grew "el Mango," a kind of rye, and "la Tuca," a kind of barley (Parodi and Hernandez 1964). Claudio Gay, who visited the province of Chiloe in 1837, collected specimens of this cereal that are currently on file at the herbarium of the Natural History Museum in Paris. He was probably the last botanist to see *B. mango* grown as a cereal inasmuch as it was replaced during the eighteenth century by

wheat and barley introduced to the New World by European settlers.

Mango was grown as a biannual. In the past, farmers allowed animals to graze on mango fields during the first year and harvested it as a cereal at the end of the next summer (Gay 1865). It is surprising that a biannual species should have been domesticated. However, it may have been the only grass in the region that lent itself to domestication. J. Ball (1884) recorded that the people of northwestern Argentina and adjacent territories harvested a species of *Bromus* as a wild cereal.

Sauwi, another native American, was extensively grown along the flood plains of the Rio Grande until the late nineteenth century (Palmer 1871; Elsasser 1979). It was sown as soon as the water receded (Kelly 1977). Today sauwi is grown only by the Warihos of the southeastern Sonora and adjacent Chihuahua of Mexico (Nabhan and de Wet 1984).

The species *Panicum sonorum* occurs as part of the natural vegetation along the western escarpment of the Sierras from southern Arizona to Honduras. It is an aggressive colonizer and often occurs in large continuous populations. It is relished by grazing animals and harvested as a wild fodder by farmers in the Mexican states of Chihuahua and Sonora. Cultivated sauwi differs conspicuously from wild *P. sonorum* in having larger spikelets that tardily disarticulate from inflorescences at maturity. Sauwi was probably domesticated by farmers who also grew other crops. It is found in an archaeological context associated with beans and cucurbits (Kaemlein 1936). This cereal is a potentially promising fodder crop in semi-arid regions of Africa and Asia.

Wild rice of North America (*Zizania aquatica* L.) is the only grass species domesticated as a cereal by present-day plant breeders. Early European explorers were impressed by the extensive use of this wild grass. In 1778, Jonathan Carver reported that wild rice was the most valuable of all the native wild food plants of the country (Carver 1778: 522-5). It had been harvested as a cereal from rivers and lakes in the northern states and adjacent Canada since long before recorded history (Coville and Coves 1894; Jenks 1900; Larsen 1939). Charred remains of wild rice caryopses found in threshing pits date from well before contact with Europeans (Ford and Brose 1975). Wild rice is now harvested from wild populations on a commercial scale, and it is also grown in paddies because such harvests cannot meet demand.

Wild rice was only domesticated very recently (de Wet and Oelke 1979). But the species does not readily lend itself to domestication: Caryopses rapidly lose viability after harvest if not stored underwater or in mud, and the species does not thrive in stagnant water. Therefore, domestication involved a combination of selection for spikelets that persisted on the panicle at maturity and the development of a crop-

ping system that took advantage of natural adaptations of the species.

Wild rice is now successfully grown on a commercial scale. Paddies are constructed so that a minimum water depth of 15 centimeters (cm) can be maintained. These are flooded and seeded in early spring. Germination is rapid, and water level is maintained until the crop matures. Fields are drained, and the crop is mechanically harvested.

African Millets

The Near Eastern cereals, wheat and barley, have been cultivated in North Africa since at least the late fifth millennium B.C. (Clark 1976). However, these temperate cereals are poorly adapted for cultivation in the tropics south of the Sahara, where eight tropical African grass species were locally domesticated.

Sorghum (*Sorghum bicolor* [L.] Moench) is the most important native African cereal. It is grown on some 50 million hectares and produces up to 80 million metric tons of grain annually, primarily for human food or animal feed. Wild sorghum is widely distributed south of the Sahara in the Sudanian climatic zone, which receives 600 to 800 millimeters (mm) of annual rainfall and extends into the wetter Guinean zone. It became a cultivated cereal at least 5,000 years ago (Connah 1967: 25; de Wet 1978; Wendorf et al. 1992).

Pearl millet (*Pennisetum glaucum* [L.] R. Br.), also called bulrush millet, is the second most important native African cereal. It is cultivated across the arid and semiarid tropics of Africa and Asia where no other cereal consistently produces a harvest because of low annual rainfall and high soil temperatures. Pearl millet is grown on about 14 million hectares in Africa and 11 million hectares in India and Pakistan. In areas where annual rainfall exceeds 600 mm in the African and Indian tropics, pearl millet is usually replaced by sorghum as a dry land crop, and in locations where annual rainfall is over 1,200 mm, finger millet or maize becomes the cereal of choice.

The species *Pennisetum glaucum* (pearl millet) is morphologically complex. O. Stapf and C. E. Hubbard (1934) recognized 13 cultivated, 15 weed, and 6 wild taxa of this complex species. W. D. Clayton (1972), by contrast, recognized 2 wild species, 2 weed taxa, and 1 cultivated complex. J. N. Brunken (1977), however, demonstrated that these taxa are conspecific. Wild taxa are often difficult to distinguish from weed taxa, which, in turn, grade morphologically into cultivated taxa. Agricultural weeds, known as shibras in West Africa, often resemble cultivated pearl millet, except for their natural seed dispersal. The involucre of cultivated pearl millet do not disarticulate at maturity. Shibras are common in the cultivated pearl millet zone of West Africa and the Sudan. They also occur in southern Angola and Namibia.

Wild pearl millet taxa occur extensively from coastal Senegal and Mauritania to northeastern Ethiopia in the Sahelo-Sudanian (350 to 600 mm annual rainfall) zone. They also extend into the Sudanian (600 to 800 mm) bioclimatic zone and are found along the foothills of mountains in the Central Sahara. Wild taxa are aggressive colonizers of disturbed habitats and are often weedy around villages.

Cultivated pearl millet is genetically and morphologically variable (Brunken 1977; Clement 1985; Marchais and Tostain 1985). Inflorescences range in shape from cylindrical to broadly elliptic and in length from 5 to 200 centimeters. Large inflorescences are commonly produced on plants with single culms, whereas small- to medium-sized inflorescences are produced on plants that tiller.

Four races of pearl millet were recognized by Brunken, J. M. J. de Wet, and J. R. Harlan (1977).

1. Race *typhoides* is grown across the range of pearl millet cultivation and is characterized by obovate caryopses that are obtuse and terete in cross section. Inflorescences are variable in length, but usually several times longer than wide, and more or less cylindrical in shape.
2. Race *nigritarum* differs from *typhoides*, primarily, in having obovate caryopses that are angular in cross section. It is the dominant pearl millet of the eastern Sahel from Sudan to Nigeria.
3. The principal pearl millet in the Sahel west of Nigeria is race *globosum*. It has large, globose caryopses, and commonly large, candle-shaped inflorescences.
4. Race *leonis* is the pearl millet common to Sierra Leone, Senegal, and Mauritania. It has oblanceolate caryopses with the apex acute.

Pearl millet could have been domesticated anywhere along the southern fringes of the Sahara (Harlan 1971). Botanical evidence indicates that the *Pennisetum violaceum* complex, as recognized by Clayton (1972), is the progenitor of domesticated pearl millet.

J. D. Clark (1976) suggested that cereal cultivation spread from the Near East to North Africa during the fifth millennium B.C. and subsequently became established across North Africa. With the onset of the present dry phase in North Africa, cultivation of these Mediterranean cereals was eventually confined to the coastal belt, and those farmers forced south by the expanding desert domesticated native grasses as cereals (Clark 1964).

Along the southern fringes of the expanding desert, the most abundant tropical grass species that invites domestication is *P. violaceum*. Its colonizing ability gives rise to large populations that facilitate its harvesting as a wild cereal. Indeed, P. J. Munson (1976) presented archaeological evidence of such harvesting along the southwestern fringes of the Sahara dating as far back as 3,000 years, and O. Davies (1968) reported

archaeological remains of cultivated pearl millet in northern Ghana dated at about the same time. Cultivated pearl millet eventually reached India as a cereal some 2,500 years ago (Rao et al. 1963).

Wild pearl millet is a typical desert grass. It produces large numbers of caryopses that can withstand heat and drought and remain dormant in the soil until conditions become favorable for germination. Caryopses germinate rapidly after the first good rains of the season, and seedlings quickly extend roots into the subsurface soil layers. Plants tiller profusely, go dormant under heat or drought stress, and produce new tillers when conditions become favorable for growth and reproduction. The strategy for survival in such a harsh environment is opportunism with respect to moisture availability and tolerance with respect to high temperature.

Cultivated pearl millet retains these adaptations. It grows and develops rapidly under conditions of adequate soil moisture and elevated temperatures, and thus can take advantage of a short growing season, to survive short periods of severe drought, and to resume growth when water becomes available again. Comparisons among genotypes indicate that differences in time of flowering under stress are the major component of yield differences among cultivars (Bidinger, Mahalakshmi, and Rao 1987). This suggests that the high degree of variability in time to flower among landrace populations is the result of natural selection for early flowering and, thus, escape from drought in dry years, and farmer selection for late flowering plants with large inflorescences in wet years (de Wet, Peacock, and Bidinger 1991). These adaptations make pearl millet the dominant cereal in the Sahelo-Sudanian zone of Africa and in semiarid regions of Zambia, Zimbabwe, Namibia, Angola, northwestern India, and adjacent Pakistan.

Finger millet (*Eleusine coracana* [L.] Gaertn.) is another native African cereal that was introduced into India during the first millennium B.C. (Vishnu-Mittre 1968). Finger millet is cultivated in wetter and cooler seasonal rainfall zones of southern Africa on about 1 million hectares and is a major cereal in the Lake Victoria region, particularly in eastern Uganda. In India, finger millet is grown on about 3 million hectares from Uttar Pradesh to Bihar and south to Tamil Nadu and Karnataka, with the states of Andhra Pradesh, Karnataka, and Tamil Nadu the major producers of this cereal. This wide distribution of finger millet has led to considerable controversy over the place of its original domestication and the identity of its wild progenitor (Hilu and de Wet 1976).

Two wild species closely resemble finger millet in gross morphology: *Eleusine indica* [L.] Gaertn., which is widely distributed in both Africa and Asia; and *E. africana* Kennedy-O'Byrne, which is predominantly African. P. J. Greenway (1945) suggested that finger millet had an African origin and that its wild progenitor is *E. africana*. But J. Kennedy-O'Byrne

(1957) proposed that *E. indica* gave rise to Indian cultivars and *E. africana* to African cultivars.

More recent cytogenetic and morphological evidence indicates that *E. africana* is the closest wild relative of finger millet. Finger millet is a tetraploid with $2n = 36$ chromosomes, as is *E. africana*, that crosses with the cereal to produce fertile hybrids. Derivatives of such crosses are obnoxious weeds of cultivation in eastern Africa. *Eleusine indica* is a diploid ($2n = 18$) and genetically isolated from the cereal.

In their work, K. W. Hilu and de Wet (1976) taxonomically recognized finger millet as *E. coracana*, subspecies *coracana*, and the wild and weedy African complex as *E. coracana*, subspecies *africana*. Wild finger millet is a common grass along the eastern and southern African highlands and is harvested during times of scarcity.

The antiquity of finger millet cultivation in eastern Africa is not known with certainty (Harlan, de Wet, and Stemler 1976). Impressions of wild, and possibly cultivated, finger millet spikelets occur on potsherds from Neolithic settlements at Kadero in Central Sudan that date back about 5,000 years (Klichowska 1984). Further archaeological evidence presented by Hilu, de Wet, and Harlan (1979) suggests that a highly evolved race of finger millet was grown at Axum, in Ethiopia, by the first century A.D. If these dates are correct, finger millet is the oldest known domesticated tropical African cereal.

This conclusion is not impossible. The concept of agriculture could have been introduced from West Asia into the Highlands of East Africa before the domestication of sorghum and pearl millet along the southern fringes of an expanding Sahara. The Near Eastern cultigens, wheat and barley, are adapted for cultivation on these highlands, and their introduction into eastern Africa could also have led to the domestication of native grasses as cereals.

Finger millet is variable in respect to inflorescence morphology, which is associated with selection and isolation of cultivars by farmers, rather than ecogeographical adaptation. Morphologically similar cultivars are widely grown, and African and Indian cultivars are often difficult to distinguish on the basis of morphology.

Five races of cultivated finger millet were recognized by de Wet and colleagues (1984b). Race *coracana* is grown across the range of finger millet cultivation in Africa and India. These cultivars resemble subspecies *africana* in having a well-developed central inflorescence branch. Inflorescence branches are 5 to 19 in number, essentially straight, and 6 to 11 cm long. In India, race *coracana* is often sown as a secondary crop in fields with pearl millet or sorghum.

The most common finger millets in both Africa and India belong to race *vulgaris*, which is also grown as a cereal in Indonesia. Inflorescence branches are straight, reflexed, or incurved, with all three types frequently occurring in the same field. In India, this race

is often planted as a dry-season crop following the harvest of irrigated rice, and in the eastern hills it is often sown in nurseries and transplanted with the first rains of the season to assure an early harvest.

With incurved inflorescence branches, race compacta resembles vulgaris cultivars, but the inflorescences are larger and the lower inflorescence branches are always divided in compacta. These cultivars are commonly known as cockscomb finger millets. Indian cultivars have a branch located some distance below the 4 to 14 main inflorescence branches, but African cultivars usually lack this lower branch. The race is grown in northeastern India, Ethiopia, and Uganda.

Race plana is grown in Ethiopia, Uganda, and the western and eastern ghats of India. Spikelets are large and arranged in two moderately even rows along the rachis, giving young inflorescence branches a ribbon-like appearance. Florets are often so numerous that they almost surround the rachis at maturity.

Race elongata is morphologically the most distinct of the five races of finger millet. Inflorescence branches are long and reflexed at maturity. Cultivars grown in Malawi have inflorescence branches up to 24 cm long. More common are cultivars with inflorescence branches of 10 to 15 cm. Race elongata is grown on the East African highlands and the hills of eastern India.

At least 1 million hectares of finger millet are planted in Africa each year. It is the preferred cereal for brewing beer and is an important food crop in Uganda, Ethiopia, Malawi, Zambia, and Zimbabwe. In India, finger millet is extensively grown by tribal people on the eastern and western ghats and by commercial farmers in Andhra Pradesh and Tamil Nadu. The area under cultivation in India is close to 3 million hectares. H. Doggett (1989) indicates that land planted with finger millet in India increased by about 3 percent annually in the 1980s. Average yield per hectare also increased from 704 kilograms (kg) in the 1950s to over 1,000 kg in the 1990s as a result of breeding African germ plasm into Indian cultivars. In East Africa a breeding program is in progress to develop cytoplasmic-genetic male sterile populations, an effort which could facilitate the production of hybrid cultivars and contribute substantially to yield increases in both Africa and India.

Tef, *Eragrostis tef* (Zucc.) Trotter is an endemic and highly valued cereal of the Ethiopian Highlands (Costanza, de Wet, and Harlan 1979). The grain is used to make *injera*, an unleavened bread that is a staple in Ethiopia, and to brew beer. The wild ancestor of tef has not yet been positively identified, but *Eragrostis pilosa* (L.) P. Beauv., a common grass on the Ethiopian highlands, is a strong possibility. T. Kotschy (1862) reported that the grains of this wild species were harvested as a food by the Sudanese while they waited for sorghum to mature. The antiquity of tef cultivation is also not known, but its popularity suggests domesti-

cation before the introduction of wheat and barley to East Africa from the Near East. W. Stiehler (1948) suggested that tef became widely distributed on the Ethiopian highlands only during the rise of the monarchy.

W. C. Harris (1844: 349) noted that 2 races of tef with brown grain and 2 with white grain were sold in Ethiopian markets. A. Trotter (1918) recognized 7 varieties of tef on the basis of spikelet and grain color.

Two species of *Digitaria* are endemic cultivated cereals of the Sahelo-Sudanian climatic zone of West Africa (Chevalier 1950; Porteres 1976). True fonio (*Digitaria exilis* [Kippist] Stapf) is grown from Senegal to Lake Chad, and black fonio (*D. iburua* Stapf) is grown on the Togo highlands and in Nigeria (de Wet 1977). The wild progenitors of these fonios are not known. In West Africa, *Digitaria barbinodis* Henrard, *D. ciliaris* Vanderyst, *D. longiflora*, and *D. ternata* (Hochst.) Stapf (Retz.) Persoon are aggressive wild colonizers and are harvested as cereals during times of scarcity.

Stapf (1915) pointed out morphological similarities between black fonio and *D. ternata*, and between fonio and *D. longiflora*. Fonio is a smaller grass than black fonio. It has 2 to 4 racemes per inflorescence, whereas black fonio has inflorescences with 4 to 10 racemes. Weedy races of fonio occur in Nigeria. Cultivated fonios differ from these weeds only in having lost their natural ability to disperse seeds efficiently.

R. Porteres (1976) recorded that fonio is harvested from some 721,000 acres annually, providing food to more than 3 million people during the most difficult months of the year. Fonios were already important in the fourteenth century when the traveler Ibn Batuta noted that they were extensively available in the markets between Outala in Mauritania and Bamako in Mali (Lewicki 1974: 37-8).

Little research has been done to improve the already impressive yield potential of fonios. Their adaptation to marginal agricultural land, tolerance to drought, and popularity as a food assure their survival as cereals in the arid Sahelian and Sudanian climatic zones of West Africa.

Animal fonio (*Brachiaria deflexa* [Schumach.] C. E. Hubb. ex Robynsis) is a weed that is commonly harvested as a wild cereal across the savanna of Africa. Farmers often encourage animal fonio to invade sorghum and maize fields, where it matures about two months before the major crop is harvested (de Wet 1977). It is sown as a cereal only on the West African Futa Jalon Highlands (Porteres 1951).

Grass weeds differ from their domesticated close relatives primarily in being spontaneous, rather than sown, and in retaining the ability of natural seed dispersal (Harlan, de Wet, and Price 1973). They do not require harvesting or sowing by humans to survive.

Avena abyssinica Hochst. is a weedy, semidomesticated of the Ethiopian highlands (Ladizinsky 1975). It is harvested, threshed, used, and sown with the wheat

or barley that it accompanies as a weed. Such cultural practices lead to a loss of natural seed dispersal ability, and as the species is not consciously sown by humans, it has become an obligate weed in cultivated fields.

Indian Millets

Wheat, rice, sorghum, pearl millet, finger millet, foxtail millet, and maize are the most important cereals grown in India. Seven indigenous cereals, mostly grown on marginal agricultural land, were domesticated in India.

Raishan (*Digitaria cruciata* [Nees] A. Camus) and adlay (*Coix lacryma-jobi* L.) are native in the wet tropics of northeastern India. Raishan is grown by the Khasi people of Assam in India and by hill tribes in Vietnam. H. B. Singh and R. K. Arora (1972) reported that this cereal is grown in Assam as a secondary crop in maize or vegetable fields. It is sown in April or May and harvested in September and October. Plants tiller profusely, and culms of individual plants are tied together at time of flowering to facilitate harvesting. Mature inflorescences are rubbed by hand to collect the grains. Dehusked grains are boiled as rice or ground into flour. Raishan is also an important fodder crop in Assam, and it could become a similarly significant fodder in other tropical regions of the world.

Adlay is grown under shifting cultivation from Assam to the Philippines (Arora 1977). It was probably domesticated in tropical eastern India and introduced into Southeast Asia, but it is also possible that adlay was independently domesticated as a cereal in both India and the Philippines. The greatest diversity of cultivated adlay occurs in the Philippines (Wester 1920).

The fruit cases of wild adlay (Job's tears) are used as beads. Fertile female spikelets of all wild *Coix* species are individually enclosed by an involucre that is indurated, glossy, and colored white, gray, or black. The involucres of cultivated adlay are papery, allowing for easy removal of the caryopses from the fruit cases. Adlay grains are cooked as rice or ground into flour to be used in baking bread. Adlay is often grown on banks between rice paddies.

The other five indigenous Indian cereals were probably domesticated in semiarid India where they still form an important part of dryland agriculture.

Sama (*Panicum sumatrense* [Roth.] ex Roem. et Schult.) is grown in India, Nepal, Sikkim, and western Myanmar (de Wet, Prasada Rao, and Brink 1984a). It is an important cereal in the eastern ghats of Andhra Pradesh and adjacent Orissa. Sama is tolerant to drought and produces a crop even in the poorest agricultural soil. It is commonly sown as a mixture with foxtail millet in sorghum or pearl millet fields, where it matures and is harvested first, followed by foxtail millet and sorghum or pearl millet. Mixed planting provides a supply of cereals starting about two

months after planting to the end of the rainy season. A robust race is sometimes planted as a single crop and is an important cereal in the hills of eastern India.

Primitive cultivars of sama resemble the widely distributed *Panicum psilopodium* Trin., except for their persistent spikelets. This wild species crosses naturally with sama to produce fertile hybrids. Derivatives of such hybrids occur as weeds in and around cultivated fields.

Sama has been grown as a cereal in India for at least 4,500 years. S. A. Weber (1991: 107–8) pointed out that carbonized grains of sama are common at the early Harappan agricultural site of Rodji.

Sawa (*Echinochloa colona* [L.] Link) is grown in India, Nepal, and Sikkim (de Wet et al. 1983a). Cultivated kinds of sawa are also known taxonomically as *E. utilis* Ohwi et Yabuno. It is morphologically allied to Japanese millet (*Echinochloa crus-galli* [L.] P. Beauv.), but sawa is tropical rather than temperate in its distribution. Furthermore, these tropical and temperate domesticated species are genetically isolated from one another (Yabuno 1966). Sawa was probably domesticated in India, whereas Japanese millet seems to have originated in northwestern China. Some Indian cultivars of sawa differ from weedy *E. colona* only in having spikelets that disarticulate tardily rather than readily at maturity, as is common in wild grasses. These weedy sawas frequently occur with cultivated races in the same field, and sawa could have been domesticated originally by an accidental harvest of weed sawas in fields where other cereals were planted.

Four races of sawa are recognized. Races have little geographic distinctiveness but are recognized and maintained by farmers. Race stolonifera resembles wild *E. colona*, except for persistence of spikelets in the cereal and disarticulation of spikelets at maturity in the weed. Race robusta has large inflorescences and is widely grown. It crosses with stolonifera. Derivatives of such hybridization gave rise to the stoloniferous race intermedia. The most distinct race is laxa. It is grown in Sikkim and is characterized by long and slender racemes.

In Africa, *Echinochloa colona* is also an aggressive colonizer of cultivated fields. D. M. Dixon (1969) identified grains of *E. colona* among plant remains from intestines of mummies excavated at Naga ed-Dar in Egypt. The species was probably harvested as a wild cereal in ancient Egypt along the flood plain of the Nile, a practice that remains common today in times of scarcity.

Kodo (*Paspalum scrobiculatum* L.) is an important cereal in Kerala and Tamil Nadu and a minor cereal in India north to Rajasthan, Uttar Pradesh, Bihar, and West Bengal. The species occurs wild across the Old World tropics (Clayton and Renvoise 1982). It is an aggressive colonizer of disturbed habitats and lends itself to domestication. Wild kodo is a perennial, whereas the cultivated cereal is grown as an annual.

Some cultivars of kodo millet root at lower nodes of their decumbent culms to produce new flowering culms after the first harvest. Kodo occurs in the agricultural record of India starting 3,000 years ago (Kajale 1977; Vishnu-Mittre 1977). Little racial evolution has occurred in Kodo millet.

The commonly grown kodo millet resembles spontaneous kinds in having racemes with spikelets arranged in two regular rows on one side of a flattened rachis. Two types of inflorescence aberrations occur occasionally in fields of kodo. In one variant, spikelets are arranged along the rachis in two to four irregular rows, rather than two regular rows. In the other variant, the spikelets are arranged in several irregular rows at the lower part of racemes and become two regular rows near the tip of the racemes. These aberrant plants are more robust and have fewer and better synchronized tillers than common kodo millet. Introgression with weed kodo makes it impossible for farmers to maintain these high-yielding genotypes, although they are carefully selected to provide seed for the next season (de Wet et al. 1983b). Farmers in southern India correctly believe that Kodo millet grains can be poisonous after a rain. The reason for this toxicity is ergot infection.

Korali (*Setaria pumila* [Poir.] Roem. et Schult.) and peda sama (*Brachiaria ramosa* [L.] Stapf) are domesticated Indian weeds that are widely distributed in tropical Africa and Asia. They are often harvested as wild cereals in times of scarcity and are cultivated only by the hill tribes of southern India. Wild and cultivated kinds of both korali and peda sama cross to produce aggressive colonizer populations. Farmers tend to keep the domesticated kinds pure through seed selection.

Eurasian Millets

Four millets – crab grass (*Digitaria sanguinalis* [L.] Scopoli), proso millet (*Panicum milliaceum* L.), foxtail millet (*Setaria italica* [L.] P. Beauv.), and Japanese millet (*Echinochloa crus-galli* [L.] P. Beauv.) – are widely distributed across temperate Europe and Asia.

Crab grass (*D. sanguinalis*) is a cosmopolitan weed. It became semidomesticated in southeastern Europe after having been harvested for millennia. Crab grass never completely lost the ability of natural seed dispersal because it was commonly harvested by swinging a basket to collect the mature grains. The species is an aggressive natural colonizer of disturbed habitats, including cultivated fields. It was a popular cereal in southern Europe during Roman times (Körnicke 1885: 279–84) and was still widely grown as mana or bluthirse in southeastern Europe during the first quarter of the nineteenth century. Crab grass was probably independently domesticated in several parts of its range. It is currently grown as a minor cereal in the Caucasus of Russia and in Kashmir.

Japanese millet (*Echinochloa crus-galli*) is a grass of temperate Eurasia. The barnyard grass found in the American Midwest is an introduced weed race of *E. crus-galli*. *Echinochloa oryzoides* (Ard.) Fritsch, the common weed of rice cultivation, is also distantly related to *E. crus-galli*. Japanese millet is grown as a cereal in China, Korea, and Japan. Cultivated kinds are sometimes, incorrectly, classified as *E. frumentacea* (Roxb.) Link. Little is known about the antiquity of this cereal. H. Helmqvist (1969) suggested that the species was grown in Sweden during the Bronze Age when the climate was milder than it is today. It is no longer grown as a cereal anywhere in Europe.

The genus *Panicum* is widely distributed throughout the warmer parts of the world and is of considerable economic importance. Several species are grown as fodder, others are harvested as wild cereals in times of scarcity, and still others are obnoxious weeds. Proso millet (*Panicum milliaceum*) was once widely cultivated in temperate Europe and Asia but has largely been replaced as a cereal by wheat.

The closest wild relative of proso millet, *Panicum milliaceum* var. *ruderales* Kitagawa is native to central China (Kitagawa 1937). In morphology, it resembles weed races of proso millet that occur across temperate Eurasia but is a less aggressive colonizer than the weed. These weeds represent derivatives of cultivated proso millet that regained the ability of natural seed dispersal through mutation (Scholz 1983).

Proso millet has been grown in central China for at least 5,000 years (Cheng 1973), and it is still grown on about 1.5 million hectares in China. It is also an important cereal in Mongolia, Korea, and northern India. A cultivar of proso millet with glutinous endosperm is favored in China, where its flour is used to make bread. Nonglutinous cultivars are grown in Mongolia and India, and the grains are cooked as rice.

Proso millet has been grown in southern Europe for at least 3,000 years (Neuweiler 1946). It became widely distributed as a cereal in Europe during the Bronze Age. Its popularity declined during the twentieth century, and proso millet is now grown in Europe primarily as a feed for caged birds. It is extensively variable. Five cultivated races are recognized. They are artifacts of selection by farmers and have no ecogeographic validity.

Race *miliaceum* resembles wild var. *ruderales* in having numerous decumbent culms, each with several racemes. Its inflorescences are large, with spreading branches that commonly lack spikelets at the base. This is the basic race from which the other races were selected under cultivation. It is grown across the range of proso millet cultivation.

Race *patentissimum* resembles *miliaceum* in its lax panicles with spreading branches having a sterile zone at the base. Inflorescences, however, become curved at maturity because of the weight of the spikelets. *Patentissimum* is the common proso millet in India, Bangladesh, Pakistan, and Afghanistan. It is

also grown in Turkey, Hungary, Russia, and China. Race *patentissimum* probably reached India from central Asia during historical times.

Races *contractum*, *compactum*, and *ovatum* are often difficult to distinguish from one another. They represent the highest evolved cultivars of proso millet. Inflorescences are more or less elliptic in shape. Spikelets are crowded along the panicle branches in *compactum* and *ovatum*. These branches are erect when young and become curved at maturity. *Ovatum* cultivars usually have smaller inflorescences than race *compactum* and are grown in Russia, Turkey, and Afghanistan. *Compactum* cultivars are grown in Japan, Russia, Iran, and Iraq. In race *contractum* the lower part of panicle branches are free of spikelets. Race *contractum* is grown in Europe, Transcaucasian Russia, and China.

Foxtail millet, *Setaria italica*, is grown as a cereal in southern Europe, in temperate Asia, and in tropical India. Its closest wild relative is the cosmopolitan weed, green foxtail (*S. italica viridis* [L.] Thell.). The latter is primarily an urban weed, but as a robust race it is also an obnoxious weed of agriculture (Pohl 1966). This giant green foxtail is derived from introgression between cultivated and wild races.

The antiquity of foxtail millet as a cereal is uncertain. The species could have been domesticated across its range of natural distribution from Europe to Japan (de Wet, Oestry-Stidd, and Cubero 1979). It has been grown as a cereal in China for at least 5,000 years (Cheng 1973) and in Europe for at least 3,000 years (Neuweiler 1946). Foxtail millet was an important cereal during the Yang-shao culture phase in China. Evidence of foxtail millet in storage jars, and the association of farming implements with the Yang-shao culture, suggest that the cereal was cultivated rather than harvested from wild populations (Ho 1975). In Europe foxtail millet commonly occurs in early farming sites in Austria and Switzerland (Neuweiler 1946).

Cultivated foxtail millets are commonly divided into two cultivated complexes. The Chinese-Korean complex with large, pendulous inflorescences is recognized as race *maxima*, and the European complex with smaller and more erect cultivars is called race *moharia* (Dekaprelevisch and Kasparian 1928). An Indian complex, race *indica*, was identified by K. E. Prasada Rao and colleagues (1987). The Indian race was derived from *moharia* through selection for adaptation to the tropics. It is an important cereal among hill tribes of southern India, where it is frequently grown as a mixed crop with sorghum or pearl millet.

F. Körnicke (1885: 238–44) recorded that canary grass (*Phalaris canariensis* L.) was grown as a cereal in southern Europe until the nineteenth century. Flour produced from its grain was mixed with wheat flour for making bread. It is still grown as a feed for birds but no longer used as a human food (Febrel and Carballido 1965). Nothing is known about the antiq-

uity of canary grass as a cereal, and it is probably of recent domestication.

J. M. J. de Wet

Bibliography

- Arora, R. K. 1977. Job's tears (*Coix lacryma-jobi*) – a minor food and fodder crop of northeastern India. *Economic Botany* 31: 358–66.
- Ball, J. 1884. Contributions to the flora of North Patagonia and the adjacent territory. *Journal of the Linnean Society of Botany, London* 21: 203–40.
- Bidinger, F. R., V. Mahalakshmi, and G. D. P. Rao. 1987. Assessment of drought resistance in pearl millet (*Pennisetum Americanum* [L.] Leeke). I. Factors affecting yields under stress. *Australian Journal of Agricultural Research* 38: 37–48.
- Brunken, J. N. 1977. A systematic study of *Pennisetum* sect. *Pennisetum* (Gramineae). *American Journal of Botany* 64: 161–7.
- Brunken, J. N., J. M. J. de Wet, and J. R. Harlan. 1977. The morphology and domestication of pearl millet. *Economic Botany* 31: 163–74.
- Callen, E. O. 1965. Food habits of some pre-Columbian Mexican Indians. *Economic Botany* 19: 335–43.
1967. The first New World cereal. *American Antiquity* 32: 535–8.
- Carver, Jonathan. 1778. *Travels through interior parts of North America in the years 1766, 1767 and 1768*. London.
- Cheng, Te-Kun. 1973. The beginning of Chinese civilization. *Antiquity* 47: 197–209.
- Chevalier, A. 1950. Sur l'origine des *Digitaria* cultives. *Revue Internationale Botanique Appliqué d'Agriculture Tropical* 12: 669–919.
- Chomko, S. A., and G. W. Crawford. 1978. Plant husbandry in prehistoric eastern North America: New evidence for its development. *American Antiquity* 43: 405–8.
- Clark, J. D. 1964. The prehistoric origins of African culture. *Journal of African History* 5: 161–83.
1976. Prehistoric populations and pressures favoring plant domestication in Africa. In *Origins of African plant domestication*, ed. J. R. Harlan, J. M. J. de Wet, and A. B. L. Stemler, 67–105. The Hague.
1984. The domestication process in northeastern Africa: Ecological change and adaptive strategies. In *Origin and early development of food-producing cultures in north-eastern Africa*, ed. L. Krzyzaniak and M. Kobusiewicz, 27–41. Poznan, Poland.
- Clayton, W. D. 1972. Gramineae. In *Flora of tropical Africa*, Vol. 3, part 2, ed. F. N. Hepper, 459–69. London.
- Clayton, W. D., and S. A. Renvoize. 1982. Gramineae. In *Flora of tropical East Africa*, part 3, ed. R. M. Polhill, 607–12. Rotterdam.
- Clement, J. C. 1985. *Les mils pénicillaires de L'Afrique de L'Ouest*. [International Board for Plant Genetic Resources (IBPGR) and Institut Français de Recherche Scientifique pour le Développement en Coopération (ORSTOM),] 1–231. Rome.
- Connah, G. 1967. Progress report on archaeological work in Bornu 1964–1966 with particular reference to the excavations at Daima mound. *Northern History Research Scheme, 2nd. Interim Report*. Zaria, Nigeria.

- Costanza, S. H., J. M. J. de Wet, and J. R. Harlan. 1979. Literature review and numerical taxonomy of *Eragrostis tef* (t'ef). *Economic Botany* 33: 413-14.
- Coville, F. V., and E. Covles. 1894. The wild rice of Minnesota. *Botanical Gazette* 19: 504-6.
- Cowan, C. W. 1978. The prehistoric use and distribution of maygrass in eastern North America: Culture and phyto-geographical implications. In *The nature and status of ethnobotany*, ed. R. I. Ford, 263-88. Ann Arbor, Mich.
- Cruz, A. W. 1972. *Bromus mango*, a disappearing plant. *Indesia* 2: 127-31.
- Davies, O. 1968. The origins of agriculture in West Africa. *Current Anthropology* 9: 479-82.
- Dekaplevich, L. L., and A. S. Kasparyan. 1928. A contribution to the study of foxtail millet (*Setaria italica* P.B. *maxima* Alf.) cultivated in Georgia (western Transcaucasia). *Bulletin of Applied Botany and Plant Breeding* 19: 533-72.
- de Wet, J. M. J. 1977. Domestication of African cereals. *African Economic History* 3: 15-32.
1978. Systematics and evolution of *Sorghum* sect. *Sorghum* (Gramineae). *American Journal of Botany* 65: 477-84.
1981. Species concepts and systematics of domesticated cereals. *Kulturpflanzen* 29: 177-98.
1989. Origin, evolution and systematics of minor cereals. In *Small millets in global agriculture*, ed. A. Seetharama, K. W. Riley, and G. Harinarayana, 19-30. Oxford and New Delhi.
- de Wet, J. M. J., and E. A. Oelke. 1979. Domestication of American wild rice (*Zizania aquatica* L., Gramineae). *Journal d'Agriculture Traditionel et Botanique Appliqué* 30: 159-68.
- de Wet, J. M. J., L. L. Oestry-Stidd, and J. I. Cubero. 1979. Origins and evolution of foxtail millets. *Journal d'Agriculture Traditionel et Botanique Appliqué* 26: 159-63.
- de Wet, J. M. J., J. M. Peacock, and F. R. Bidingier. 1991. Adaptation of pearl millet to arid lands. In *Desertified grasslands: Their biology and management*, ed. G. F. Chapman, 259-67. London.
- de Wet, J. M. J., K. E. Prasada Rao, and D. E. Brink. 1984a. Systematics and domestication of *Panicum sumatrense* (Gramineae). *Journal d'Agriculture Traditionel et Botanique Appliqué* 30: 159-68.
- de Wet, J. M. J., K. E. Prasada Rao, D. E. Brink, and M. H. Mengesha. 1984b. Systematics and evolution of *Eleusine coracana* (Gramineae). *American Journal of Botany* 71: 550-7.
- de Wet, J. M. J., K. E. Prasada Rao, M. H. Mengesha, and D. E. Brink. 1983a. Domestication of sawa millet (*Echinochloa colona*). *Economic Botany* 37: 283-91.
- 1983b. Diversity in Kodo millet, *Paspalum scrobiculatum*. *Economic Botany* 37: 159-63.
- Dixon, D. M. 1969. A note on cereals in ancient Egypt. In *The domestication and exploitation of plants and animals*, ed. J. P. Ucko and C. W. Dingleby, 131-42. Chicago.
- Doggett, H. 1989. Small millets - a selective overview. In *Small millets in global agriculture*, ed. E. Seetharama, K. W. Riley, and G. Harinarayana, 3-17. Oxford and New Delhi.
- Elsasser, A. B. 1979. Explorations of Hernando Alcaron in the lower Colorado River, 1540. *Journal of California Great Basin Anthropology* 1: 8-39.
- Febrel, J., and A. Carballido. 1965. Estudio bromatológico del alpieste. *Estudio Bromatológico* 17: 345-60.
- Ford, R. I., and D. S. Brose. 1975. Prehistoric wild rice from Dunn farm site, Leelanau Country, Michigan. *The Wisconsin Archaeologist* 56: 9-15.
- Gay, C. 1865. *Historia física y política de Chile*, Vol. 6. Reprinted in *Agricultura Chilena*, Vol. 2, 1973. Santiago.
- Greenway, P. J. 1945. Origin of some East African food plants. *East African Agricultural Journal* 10: 177-80.
- Harlan, J. R. 1971. Agricultural origins: Centers and non-centers. *Science* 174: 463-74.
- Harlan, J. R., J. M. J. de Wet, and E. G. Price. 1973. Comparative evolution of cereals. *Evolution* 27: 311-25.
- Harlan, J. R., J. M. J. de Wet, and A. B. L. Stemler. 1976. Plant domestication and indigenous African agriculture. In *Origins of African plant domestication*, ed. J. R. Harlan, J. M. J. de Wet, and A. B. L. Stemler, 3-19. The Hague.
- Harris, W. C. 1844. *The highlands of Aethiopia*. New York.
- Helmqvist, H. 1969. Dinkel und Hirse aus der Bronzezeit Südschwedens nebst einigen Bemerkungen über ihre spätere Geschichte in Sweden. *Botanische Notizen* 122: 260-70.
- Hilu, K. W., and J. M. J. de Wet. 1976. Domestication of *Eleusine coracana*. *Economic Botany* 306: 199-208.
- Hilu, K. W., J. M. J. de Wet, and J. R. Harlan. 1979. Archaeobotanical studies of *Eleusine coracana* ssp. *coracana* (finger millet). *American Journal of Botany* 66: 330-3.
- Ho, Ping-Ti. 1975. *The cradle of the east*. Chicago.
- Jenks, A. E. 1900. The wild rice gatherers of the upper lakes. *Annual Report of the American Bureau of Ethnology* 1989, part 2, 19: 1013-137.
- Kaemlein, W. 1936. A prehistoric twined-woven bag from the Trigo Mountains, Arizona. *Kiva* 28: 1-13.
- Kajale, M. P. 1977. Ancient grains from excavations at Nevassa, Maharashtra. *Geophytologia* 7: 98-106.
- Kelly, W. H. 1977. Cocopa ethnography. *Anthropological Papers of the University of Arizona* 29: 1-150.
- Kennedy-O'Byrne, J. 1957. Notes on African grasses. 24. A new species of *Eleusine* from tropical and South Africa. *Kew Bulletin* 11: 65-72.
- Kitagawa, M. 1937. Contributio ad cognitionem Florae Manchuricae. *Botanical Magazine of Tokyo* 51: 150-7.
- Klichowska, M. 1984. Plants of the Neolithic Kadero (central Sudan): A palaeobotanical study of the plant impressions on pottery. In *Origins and early development of food-producing cultures in north-eastern Africa*, ed. L. Krzysniak and M. Kobusiewicz, 321-6. Poznan, Poland.
- Körnicker, F. 1885. Die Arten und Varietäten des Getreides. In *Handbuch des Getreidebaues*, ed. F. Körnicker, and H. Werner, Vol. 1. Berlin.
- Kotschy, T. 1862. Reise von Chartum nach Kordafan, 1839. *Petermann's geographische Mittheilungen Ergänzungsbeft* 7: 3-17.
- Ladizinsky, G. 1975. Oats in Ethiopia. *Economic Botany* 29: 238-41.
- Larsen, E. L. 1939. Peter Kalm's short account of the natural use and care of some plants, of which the seeds were recently brought home from North America to the service of those who take pleasure in experimenting with the cultivation of the same in our climate. *Agricultural History* 13 (34): 43-4.
- Lewicki, T. 1974. *West African food in the Middle Ages*. London.
- Marchais, L., and S. Tostain. 1985. Genetic divergence between wild and cultivated pearl millets (*Pennisetum typhoides*). II. Characters of domestication. *Zeitschrift für Pflanzenzüchtung* 95: 245-61.

- Munson, P. J. 1976. Archaeological data on the origins of cultivation in the southwestern Sahara and their implications for West Africa. In *Origins of African plant domestication*, ed. J. R. Harlan, J. M. J. de Wet, and A. B. L. Stemler, 187-209. The Hague.
- Nabhan, G., and J. M. J. de Wet. 1984. *Panicum sonorum* in the Sonoran desert agriculture. *Economic Botany* 38: 65-82.
- Neuweiler, E. 1946. Nachträge urgeschichtlicher Pflanzen. *Vierteljährliche Naturforschungsgesellschaft Zürich* 91: 122-236.
- Palmer, E. 1871. Food products of North American Indians. *United States of America Commerce and Agricultural Report* 1870: 404-28.
- Parodi, L. R., and J. C. Hernandez. 1964. El mango, cereal extinguido en cultivo, sobre en estado salvaje. *Ciencia e Invéstita*: 20: 543-9.
- Phillips, S. M. 1972. A survey of the genus *Eleusine* Gaertn. (Gramineae) in Africa. *Kew Bulletin* 27: 251-70.
- Pohl, R. W. 1966. The grasses of Iowa. *Iowa State Journal of Science* 40: 341-73.
- Porteres, R. 1951. Une céréale mineure cultivée dans l'Ouest-Afrique (*Brachiaria deflexa* C.E. Hubbard var. *sativa* var. nov.). *L'Agronomie Tropicale* 6: 38-42.
1976. African cereals: Eleusine, fonios, black fonio, teff, brachiaria, paspalum, pennisetum and African rice. In *African plant domestication*, ed. J. R. Harlan, J. M. J. de Wet, and A. B. L. Stemler, 409-52. The Hague.
- Prasada Rao, K. E., J. M. J. de Wet, D. E. Brink, and M. H. Mengesha. 1987. Intraspecific variation and systematics of cultivated *Setaria italica*, foxtail millet (Poaceae). *Economic Botany* 41: 108-16.
- Rao, S. R., B. B. Lal, B. Nath, et al. 1963. Excavations at Rangpur and other explorations in Gujarat. *Bulletin of the Archaeological Survey of India* 18-19: 5-207.
- Scholz, H. 1983. Die Unkraut-Hirse (*Panicum miliaceum* subsp. *ruderales*) - neue Tatsachen und Befunde. *Plant Systematics and Evolution* 143: 233-44.
- Singh, H. B., and R. K. Arora. 1972. Raishan (*Digitaria* sp.) - a minor millet of the Khasi Hills, India. *Economic Botany* 26: 376-90.
- Stapf, O. 1915. Iburu and fundi, two cereals of Upper Guinea. *Kew Bulletin* 1915: 381-6.
- Stapf, O., and C. E. Hubbard. 1934. *Pennisetum*. In *Flora of tropical Africa*, Vol. 9, ed. D. Prain, 954-1070. London.
- Stiehler, W. 1948. Studien zur Landwirtschafts und Siedlungsgeographie Äthiopiens. *Erdkunde* 2: 257-82.
- Trotter, A. 1918. La *Poa tef* Zuccagni e l'*Eragrostis abyssinica* (Jacq.) Link. *Bolletino della Società Botanica Italiana* 4: 61-2.
- Vishnu-Mittre. 1968. Prehistoric records of agriculture in India. *Transactions of the Bose Research Institute* 31: 87-106.
1977. Changing economy in ancient India. In *Origins of agriculture*, ed. C. A. Reed, 569-88. The Hague.
- Weber, S. A. 1991. *Plants and Harappan subsistence*. Delhi.
- Wendorf, F., A. E. Close, R. Schild, et al. 1992. Saharan exploitation of plants 8000 B.P. *Nature* 359: 721-4.
- Wester, P. S. 1920. Notes on adlay. *Philippine Agricultural Review* 13: 217-22.
- Wills, W. H. 1988. Early agriculture and sedentism in the American Southwest and interpretations. *Journal of World History* 2: 445-88.
- Yabuno, T. 1966. Biosystematics of the genus *Echinochloa* (Gramineae). *Japanese Journal of Botany* 19: 277-323.

II.A.6 Oat

Oat (*Avena* L.) includes 29 to 31 species (depending on the classification scheme) of wild and domesticated annual grasses in the family Gramineae (Poaceae) that comprise a polyploid series, with diploid, hexaploid, and tetraploid forms (Baum 1977; Leggett 1992). The primary cultivated species are hexaploids, *A. sativa* L. and *A. byzantina* C. Koch, although 5 other species have to some extent been cultivated for human consumption. These are the tetraploid *A. abyssinica* Hochst and the diploids *A. strigosa* Schreb., *A. brevis* Roth., *A. hispanica* Ard., and *A. nuda* L. Nevertheless, oat consumed in human diets this century has been almost exclusively hexaploids.

The separation of the two cultivated hexaploids is based on minor, and not always definitive, morphological differences and is of more historical than contemporary relevance. *A. byzantina* (red oat) was the original germ plasm base of most North American fall-sown cultivars, whereas *A. sativa* was the germ plasm base of spring-sown cultivars. Late twentieth-century breeding populations in both ecogeographic regions contain intercrosses of both species. This has led to the almost exclusive use of the term *A. sativa* in describing new cultivar releases.

Oat



Oat is the fifth most economically important cereal in world production after wheat, rice, corn, and barley. It is cultivated in temperate regions worldwide, especially those of North America and Europe, where it is well adapted to climatic conditions of adequate rainfall, relatively cool temperatures, and long days (Sorrells and Simmons 1992).

Oat is used primarily for animal feed, although human consumption has increased in recent years. Human food use is estimated at 16 percent of total world production and, among cereals as foods, oat ranks fourth after wheat, rice, and corn (Burnette et al. 1992). Oat is valued as a nutritious grain; it has a high-quality protein and a high concentration of soluble fiber, and is a good source of minerals, essential fatty acids, B vitamins, and vitamin E (Peterson 1992). Humans generally consume oats as whole-grain foods, which include ready-to-eat breakfast foods, oatmeal, baked goods, infant foods, and granola-type snack bars (Burnette et al. 1992). For human food, the inedible hull (lemma and palea) must be removed, leaving the groat for further processing. A hull-less trait, governed by two or three genes, causes the caryopsis or groat to thresh free of the lemma and palea, as does the wheat caryopsis. There is a renewed interest in hull-less oat in Europe and the Americas, especially for feed use. Several modern hull-less cultivars are available.

Origin and Domestication

The history of oat domestication parallels that of barley (*Hordeum vulgare* L.) and wheat (*Triticum* spp.), the primary domesticated cereals of the Middle East. The primacy of wheat and barley in the Neolithic revolution was due to advantages that their progenitor species had over other local candidates, such as *A. sterilis* L. and *A. longiglumis* Dur.: local abundance, large seed weight and volume, absence of germination inhibitors, and lower ploidy levels (Bar-Yosef and Kislev 1989). In the archaeological record, wild oat appears as weedy admixtures in cultivated cereals prior to, and for several millennia following, the Neolithic revolution. Nondomesticated *Avena* spp. have been identified in archaeological deposits in Greece, Israel, Jordan, Syria, Turkey, and Iran, all dating from between about 10500 and 5000 B.C. (Hopf 1969; Renfrew 1969; Hansen and Renfrew 1978; Hillman, Colledge, and Harris 1989).

Wheat and barley remained predominant as cereal cultivation spread through Europe between the seventh and second millennium B.C. (Zohary and Hopf 1988). The precise time and location of the domestication of oat from the weedy component of these cereals is unknown, but it is believed that oat had an adaptive advantage (over the wheat and barley germ plasm in cultivation at that time) in the cloudier, wetter, and cooler environments of northern Europe.

Support for this theory is provided by Pliny (A.D.

23–79), who noted the aggressive nature of weed oat in cereal mixtures in moist environments (Rackham 1950). Z. V. Yanushevich (1989) reported finding *Avena* spp. in Moldavian and Ukrainian adobe imprints dated as early as 4700 B.C. It is not known if these were cultivated types. However, Z. Tempir, M. Villaret-von Rochow (1971), and U. Willerding (Tempir and the latter are cited in Zohary and Hopf 1988), working in central Europe, found evidence of domesticated oat dating from the second and first millennia B.C. That evidence (which is often a reflection of one of the first steps in the domestication of oat and other cereals) is the elimination of the seed dispersal mechanism. In domesticated oat, the spikelets remain intact on the plant long after ripeness, whereas in the wild species, spikelets abscise and fall from the plant soon after maturity (Ladizinsky 1988).

In China, oat has been cultivated since early in the first millennium A.D. It remains a staple food in north China and Mongolia (Baum 1977), and the hull-less or “naked” oat has been associated with the Chinese production. But despite the cultivation of oat elsewhere, the grain was of minor interest to the Greeks, who considered it a weed; moreover, Egyptian foods are not known to have contained oat and, unlike so many other foods, there is no reference to it in the Bible (Candolle 1886; Darby, Ghalioungui, and Grivetti 1977; Zohary 1982).

During the first century A.D., however, Roman writers began making references to oat (White 1970). Pliny described a fall-sown, nonshattering “Greek-oat” used in forage production and noted that oatmeal porridge was a human staple in Germany. Dioscorides described the medicinal qualities of oat and reported it to be a natural food for horses (Font Quer 1962). Analyses of the gut contents of a mummified body from the same era (recovered from an English bog) revealed that small quantities of *Avena* species, together with wheat and barley, were consumed in the final meal (Holden 1986).

J. R. Harlan (1977) believed that oat domestication occurred separately at each ploidy level, with the diploids domesticated primarily as fodder crops in the Mediterranean area and subsequently widely cultivated throughout northern and eastern Europe, particularly on poor, upland soils. The tetraploid *A. abyssinica*, found exclusively in the highlands of Ethiopia, is an intermediate form between a truly wild and a fully domesticated type. It is a tolerated admixture in cereal production because of the belief that it improves the quality of malt (Harlan 1989). There is, however, disagreement among researchers as to the hexaploid progenitor of cultivated hexaploid oat. Three species – *A. sterilis* L. (Coffman 1946; Baum 1977), *A. hybrida* Petrem. (Baum 1977), and *A. fatua* L. (Ladizinsky 1988) – have been suggested. *A. maroccana* Gdgr. and *A. murphyi* Ladiz. are believed to represent the tetraploid base for cultivated hexaploids (Ladizinsky 1969). These two species have a narrow

geographic distribution, confined as they are to southern Spain and Morocco.

From the Roman Era to the Nineteenth Century

Since the Roman era, oat has maintained its dual role as food and feed. In overall European production it has ranked behind wheat, barley, and, in some areas, rye (*Secale cereale* L.). Its prominence in the human diet continued to be greater in northern Europe than in other regions.

During the Middle Ages in northern Europe, a typical three-year crop rotation was fallow followed by wheat followed by oat or barley (Symon 1959). P. D. A. Harvey (1965) provided detailed records from one English village; these most likely reflected usual thirteenth-century production practices. Wheat was planted on half of the total cereal hectareage, with oat planted on one-half to three-quarters of the remaining land. The yield of the oat crop was about five seeds per seed planted, very low by today's standards. Wheat was the cash crop, whereas the oat crop was employed on the farms for feeding horses and cattle; in addition, oat straw was likely an important animal bedding. Lastly, another significant hectareage was planted with barley and oat together, and a small quantity of this mixed crop was used to produce malt. It is interesting to note that Arthur Young (1892) reported similar production practices in Ireland during the mid-eighteenth century.

Oat in the Western Hemisphere

Oat, both wild and cultivated, entered the Americas by two routes (Coffman 1977). *A. byzantina*, as well as the wild-weedy *A. fatua* L. and *A. barbata* Pott ex Link, were all introduced to southern latitudes by the Spaniards. *A. sativa* (and probably *A. fatua*) was transported by the English and other Europeans to the northern colonies. In seventeenth-century colonial agriculture, oat grain was fed to horses and mixed with rye and pea (*Pisum arvense* L.) for cattle feed (Bidwell and Falconer 1925). Where Scottish colonists predominated, it sometimes contributed to the human diet, although oat was not widely grown in the southern colonies of British North America (Grey and Thompson 1933). E. L. Sturtevant (1919) noted that Native Americans in California gathered wild oat and used it in breadmaking.

Oat cultivation moved west with frontier farming. Typical pioneer farmers planted maize (*Zea mays* L.) and some potatoes in the newly turned sod; this was followed the second year with small grains (Bidwell and Falconer 1925). Oat yields of 70 bushels per acre (3,760 kilograms per hectare) were achieved in Indiana by 1838 (Ellsworth 1838). To the north, in Canada, nineteenth-century pioneers relied less on maize and more on wheat and oat.

During the twentieth century, oat production in North America has been concentrated in the north central United States and the prairie provinces of

Canada. Spring-sown oat is grown in these areas, whereas fall-sown, or "winter," oat is grown in the southern and southwestern United States (and parts of Europe). Fall sowing permits the crop to grow during mild weather in late autumn and early spring and to mature prior to the onset of high summer temperatures. Fall-sown oat is grazed for winter forage in the southwestern United States. The northern limits of fall-sown production are Kentucky, southern Pennsylvania, and Virginia. Although oat thrives in cooler climates than do wheat and barley, it is more at risk from freezing temperatures than those cereals.

The foundation germ plasm for spring-sown oat in the United States and Canada consists of three Russian cultivars (Kherson, Green Russian, and White Russian), a Swedish cultivar (Victory), and a Greek cultivar (Markton). All five heterogeneous cultivars were introduced into North America during the late nineteenth or early twentieth centuries (Coffman 1977).

The foundation germ plasm for fall-sown oat in the United States comprises two heterogeneous cultivars, Winter Turf and Red Rustproof. Winter Turf was probably introduced from northern Europe, and it may have been cultivated since colonial times (Coffman 1977). Red Rustproof was introduced from Mexico and became very popular after the Civil War. Cultivars resulting from selections within this heterogeneous germ plasm dominated production from Virginia to Texas and California until well into the twentieth century. Fall-sown oat in the United States is used as feed and does not contribute to the human diet. This is not a reflection of its nutritional value; rather, it reflects the proximity of processing mills to the centers of spring oat production in the north central United States and Canada. Oat grain has a relatively low density, which makes transportation expensive.

Progress from Plant Breeding

Oat cultivar improvement through plant selection began in Europe approximately 200 years ago, but, prior to the late nineteenth century, the intensity of the effort remained small by modern standards. In parallel with the development of other cereal grains, oat breeding activity increased worldwide during the early twentieth century. Oat breeding has remained a public sector activity, with a few notable exceptions, but some of these public sector programs in North America are dependent on the financial support of private sector endusers in the food industry. Comprehensive histories of oat breeding have been produced by T. R. Stanton (1936); F. A. Coffman, H. C. Murphy, and W. H. Chapman (1961); and M. S. McMullen and E. L. Patterson (1992).

The progression of methodologies utilized in oat improvement has been similar worldwide. The initial method was the introduction of heterogeneous cultivars from one production region to another for direct

cultivation. This was gradually replaced by the development of cultivars from plant selections made within these heterogeneous introductions now growing in a new environment. The third step in the progression was the selection of cultivars from within breeding populations developed by sexual hybridization between parents with complementary arrays of desirable traits. In general, the end product of such a program was a homogeneous pure line cultivar. In the United States, the era of introduction lasted from colonial times to the beginning of the twentieth century. The era of selection within introductions as the predominant source of new cultivars extended from approximately 1900 to 1930. Since that time, the majority of cultivars have resulted from hybridization.

Common themes in oat breeding research during the twentieth century have included field, laboratory, and greenhouse methodologies to improve efficiency of selection for an array of agronomic, disease- and insect resistance, morphologic, and grain-quality traits. In addition, there have been Mendelian and quantitative genetic studies to investigate inheritance and expected progress from selection for these traits; studies of the evolution of species within the genus *Avena*; and, recently, the use of nonconventional techniques centered around biotechnology. This body of work has provided extensive knowledge of the basic biology of oat and has fostered direction and efficiency in applied oat improvement.

Throughout the twentieth century, breeding efforts have been directed at the improvement of grain yield, straw strength, test weight, and resistance to disease and insect pests. Additional efforts have been directed towards groat percentage, kernel weight, and winter hardiness. The cumulative results of these breeding efforts have included notable improvements in all of these areas (Lawes 1977; Rodgers, Murphy, and Frey 1983; Wych and Stuthman 1983; Marshall 1992; Lynch and Frey 1993), as well as the maintenance of disease and insect resistance levels. Yield improvements were not associated with specific phenotypic characteristics (as, for example, with reduced-height genes in wheat) or with germ plasm source, but modern cultivars appeared more adapted, both to high productivity and to heat- and drought-stressed environments, than older cultivars. R. D. Wych and D. D. Stuthman (1983) reported increases in biomass, total N, groat N, and nitrogen harvest index. In general, tillers per plant and kernels per panicle either have remained unchanged or have been reduced. The importance of increased biomass has been emphasized as a route to further yield increases (Moser and Frey 1994). This biomass must result from improved growth rate rather than extended growth duration, and harvest index must be maintained at present levels (Takeda and Frey 1976; Reysack, Stuthman, and Stucker 1993).

Much effort has been directed toward the identification and utilization of novel sources of disease resis-

tance in cultivar development. Crown rust (*Puccinia coronata* Cda. var. *avenae* Fraser and Led.), stem rust (*P. graminis* Pers. f. sp. *avenae* Ericks. and E. Henn.), loose smut (*Ustilago avenae* [Pers.] Rostr.), powdery mildew (*Erysiphe graminis* DC. f. sp. *avenae* Em. Marchal), and barley yellow dwarf virus have received the most attention. For several decades, breeders have been utilizing the wild hexaploid *A. sterilis* as a source of genes for protection against crown rust and other pathogens. Other, more distantly related, species have been utilized to a lesser extent (Sharma and Forsberg 1977; Aung and Thomas 1978). Multiline oat cultivars were developed in the midwestern United States as an alternative strategy for crown rust control (Frey, Browning, and Simons 1985). A multiline cultivar is a mixture of several phenotypically similar genotypes, but each genotype contains a different gene for crown rust resistance. Multilines differ from most late-twentieth-century oat cultivars in that they are not homogeneous pure lines.

Breeding for improved grain composition – that is, groat protein, groat oil, and beta-glucan content – has been emphasized during the past 25 years. Although test weight is the primary quality factor used in purchasing oat, high-yielding cultivars with elevated groat protein levels have been released with regularity during the past 20 years. The range of groat protein in these cultivars is 18 to 21 percent versus 14 to 17 percent in conventional cultivars. The impetus behind this work is the enhancement of the feed value of oat, the maintenance of its standing as a traditional breakfast food, and the increase of its potential for use in the specialty food market (for example, as a protein additive). It is noteworthy that because of the predominance of the globulin fraction in oat storage protein, oat protein quality does not decrease with increases in groat protein percentage (Peterson 1976).

Although an overall negative association between grain yield and groat protein percentage is found in oat, studies consistently report the occurrence of high-protein transgressive segregates with overall agronomic superiority. When breeders have used protein yield (grain yield $\times$ groat protein concentration) as the unit of selection, they have been effective in improving both traits simultaneously (Kuenzel and Frey 1985; McFerson and Frey 1991).

Among other findings of importance to the improvement of oat protein are that groat protein is polygenically inherited, but heritability levels are moderate; that gene action is primarily additive; and that genes from *A. sativa* and *A. sterilis* may act in a complementary fashion (Campbell and Frey 1972; Iwig and Ohm 1976; Cox and Frey 1985). Breeders have been directing most of their efforts to the wild *A. sterilis* species as a source of genes with which to increase groat protein percentage and protein yield. Other species, such as *A. fatua* and *A. magna*, have been identified as potentially valuable resources as well (Thomas, Haki, and Arangzeb 1980; Reich and

Brinkman 1984). Oat researchers believe that a further 4 to 5 percent increase in groat protein over that of current high-protein cultivars is a reasonably obtainable breeding objective.

Groat oil content in cultivars typically ranges between 3.8 and 11 percent (Hutchinson and Martin 1955; Brown, Alexander, and Carmer 1966). Approximately 80 percent of the lipid in the groat is free lipid (ether extracted), and triglycerides are the most abundant component of oat oil. Most of the total lipid is found in the bran and starchy endosperm (Youngs, Püskülcü, and Smith 1977). Oat has never been utilized as an oilseed crop, but its mean lipid content is higher than that of other temperate cereal grains. No oat cultivar has been released based solely on elevated oil content, but considerable effort has been expended on studies of this trait, with the goal of alteration through breeding. Initial interest was related to the improvement of the energy value of oat as a livestock feed. Subsequently, V. L. Youngs, M. Püskülcü, and R. R. Smith (1977) indicated that high groat oil concentration could also increase food caloric production, and K. J. Frey and E. G. Hammond (1975) estimated that oat cultivars with 17 percent groat oil (combined with present levels of protein and grain yield) would compete with Iowa soybeans as an oilseed crop for the production of culinary oil.

Inheritance of oil content was studied in crosses of cultivated oat with *A. sterilis* and *A. fatua*, and results indicated that oil content was polygenically inherited, that additive gene effects predominated, that environmental influences were minor, that transgressive segregation was common, and that heritability was high (Baker and McKenzie 1972; Frey, Hammond, and Lawrence 1975; Luby and Stuthman 1983; Thro and Frey 1985; Schipper and Frey 1991a). Thus, when a concerted effort was made to improve groat oil content and oil yield (grain yield $\times$ groat oil concentration), the results were impressive (Branson and Frey 1989; Schipper and Frey 1991b). Agronomically desirable lines with up to 15 percent groat oil content were developed rather rapidly. Lower seed and test weights were associated with high groat oil content. Subsequently, lines with oil as high as 18 percent were produced (K. J. Frey, personal communication).

The major fatty acids of oat are palmitic (16:0), stearic (18:0), oleic (18:1), linoleic (18:2), and linolenic (18:3). Of these, palmitic, oleic, and linoleic constitute 95 percent of the fatty acids measured. Oleic and linoleic are comparable in quantity and may be controlled by the same genetic system (Karow and Forsberg 1985). Increased lipid content is correlated with an increase in oleic acid and a decrease in palmitic, linoleic, and linolenic acids (Forsberg, Youngs, and Shands 1974; Frey and Hammond 1975; Youngs and Püskülcü 1976; Roche, Burrows, and McKenzie 1977).

The oil content of advanced breeding lines is monitored routinely by breeders, but fatty acid content is not usually determined. Both simple and polygenic

inheritance is involved in the expression of fatty acid content, but heritabilities are moderate to high (Thro, Frey, and Hammond 1983; Karow and Forsberg 1984). Selection for increased oil content should be accompanied by the monitoring of fatty acid composition, with particular attention to palmitic and linoleic acid, if conservation of the fatty acid composition of oat is desired.

Oat genotypes range in beta-glucan concentration from about 2.5 to 8.5 percent (D. M. Peterson and D. M. Wesenberg, unpublished data), but the range in adapted genotypes is narrower (Peterson 1991; Peterson, Wesenberg, and Burrup 1995). Several plant breeders have begun to make crosses with high beta-glucan germ plasm in an attempt to develop cultivars specially suited for human food. High beta-glucan oats are unsuited for certain animal feeds, especially for young poultry (Schrickel, Burrows, and Ingemansen 1992).

World Oat Production

The countries of the Former Soviet Union (FSU), North America, and Europe account for 90 percent of the world's oat production (Table II.A.6.1). Australia produces 4 percent and the People's Republic of China less than 2 percent.

The highest yields are obtained in the United Kingdom, Denmark, Germany, France, the former Czechoslovakia, New Zealand, and Sweden. Cool, moist summers, combined with intensive management practices, are commonplace in these countries. Large-scale producers, such as the United States, Canada, and the FSU, sacrifice high yield per hectare for less intensive management practices. Oat is adapted to cool, moist environments and is sensitive to high temperatures from panicle emergence to physiological maturity. It is more tolerant of acid soils than are other small grains, but less so of sandy or limestone soils. Although oat is adapted to cool temperatures, it is not as winter hardy as wheat or barley. Thus, the bulk of the world's production comes from spring-sown cultivars.

Of the major producers, only the FSU increased production and hectareage over the past three decades. Expanding livestock numbers coupled with the more favorable growing environment in northern regions have made oat more attractive than wheat or barley. The FSU now accounts for 40 percent of world production. All other major producers, including Canada, the United States, Germany, and Poland, have had declining production and hectareage during the same period. Production has declined by 33 percent in Poland and 61 percent in the United States and France. Overall, world production has declined by 23 percent and hectareage by 27 percent, whereas yield per hectare has increased 6 percent over the past 30 years. But production did increase in Australia, New Zealand, South America, Mexico, and Africa during the same period.

The reasons for the generally downward trend in

Table II.A.6.1. *World oat production, area harvested, and yield by continent and country, 1965 through 1994. Continent totals may not sum to world total due to rounding*

	Year and mean production			Year and area harvested			Year and mean yield		
	1965-74	1975-84	1985-94	1965-74	1975-84	1985-94	1965-74	1975-84	1985-94
	1000 t			1000 ha			t ha ⁻¹		
North America									
Canada	5203	3542	2968	2878	1782	1280	1.82	1.99	2.31
Mexico	46	78	100	61	84	101	0.78	0.92	1.00
U.S.	11884	8058	4645	6567	4237	2322	1.81	1.91	2.00
Total	17133	11678	7713	9506	6103	3703	-	-	-
South and Central America									
Argentina	491	534	474	380	392	381	1.29	1.36	1.23
Chile	116	138	177	80	84	68	1.45	1.63	2.60
Uruguay	61	40	50	79	49	56	0.76	0.82	0.89
Other	71	149	284	88	154	266	0.81	0.97	1.07
Total	739	861	985	627	679	771	-	-	-
Asia									
P.R. China	819	743	644	1050	710	562	0.78	1.05	1.15
Japan	76	16	6	36	7	3	2.13	2.36	1.88
Turkey	432	355	290	339	208	154	1.29	1.73	1.89
Total	1327	1114	940	1425	925	719	-	-	-
Africa									
Algeria	36	75	68	57	114	110	0.63	0.71	0.64
Morocco	19	31	46	24	39	49	0.85	0.76	0.92
S. Africa	107	78	54	587	456	663	0.18	0.18	0.09
Total	162	184	168	668	609	822	-	-	-
Oceania									
Australia	1233	1363	1576	1343	1205	1103	0.91	1.12	1.41
New Zealand	49	59	72	17	16	18	2.95	3.61	3.80
Total	1282	1422	1648	1360	1221	1121	-	-	-
Former Soviet Union									
	11088	14404	15221	9447	12167	11042	1.16	1.19	1.38
Western Europe									
Denmark	706	202	131	189	57	30	3.74	3.72	4.54
Finland	1133	1272	1158	502	467	382	2.25	3.00	3.05
France	2414	1812	947	877	550	240	2.80	3.34	3.97
Germany	3590	3781	2356	1060	997	533	3.39	3.79	4.26
Sweden	1468	1403	1355	477	452	380	3.07	3.39	3.59
U.K.	1213	637	525	354	162	107	3.46	4.06	4.93
Other	2523	2449	1968	1497	1238	727	1.69	1.98	2.71
Total	13047	11556	8440	4956	3623	2399	-	-	-
Eastern Europe									
Czechoslovakia	801	458	364	356	160	100	2.30	2.93	3.63
Poland	2981	2541	1993	1350	1084	779	2.22	2.35	2.53
Former Yugoslavia	322	294	233	280	204	129	1.15	1.46	1.80
Other	320	258	473	276	168	268	1.16	1.54	1.76
Total	4424	3551	3063	2262	1616	1276	-	-	-
World Total	49187	44699	38082	30195	27166	22002	1.63	1.65	1.73

Source: USDA, Economic Research Service, Washington, D.C.

oat production have included competition from crops that produce higher levels of energy and protein (such as maize and soybeans), the decline of oat use as a feed grain, changing crop rotation patterns, and government commodity programs that are more favorable to the growing of other crops. Although 79 percent of

the crop is used for feed worldwide, changes in production and use in such countries as the United States have resulted in up to 42 percent of the crop going for food and seed in recent years. Over the past decade, the United States has imported an amount equivalent to 14 percent of its annual production.

Oat Milling

The milling of oat for human food typically involves several steps: cleaning, drying, grading, dehulling, steaming, and flaking. In addition, a cutting step may be inserted after dehulling (Deane and Commers 1986). The purpose of oat milling is to clean the grain, remove the inedible hull, and render the groat stable and capable of being cooked in a reasonable time. The history of oat usage for human food is associated with the development of milling technology, which evolved slowly over the millennia and more rapidly over the past two centuries. Most of the early advancements in oat milling were ancillary to improvements in wheat milling.

Primitive peoples prepared oat by crushing the grains between two rocks. As the respective rocks wore into an oval and cup shape, a mortar and pestle were developed. This evolved into the saddlestone, where the grain was ground in a saddlelike depression by the forward and back action of an oval stone. The next development was the quern, which appeared, according to R. Bennett and J. Elton (1898), about 200 B.C.

The quern was a distinct advancement, in that the action involved a rotating stone and a stationary one, rather than an oscillatory movement. The rotating stone typically had a handle for applying the motive force and a hole in the center through which the grain was fed. Further developments included the grooving of the flat surfaces to provide a cutting edge and a channel for the flour, groats, and hulls to be expelled (Thornton 1933). In more sophisticated mills, additional stones were used – one to remove the hull and a second to crush the groat (Lockhart 1983).

Over the next 1,500 years or so, the principal advancements were in the power source, evolving from human-powered to animal-powered, and later to the use of water and wind to turn the stone. In Scotland, the first water-powered mills were in existence by the eleventh century (Lockhart 1983). By the late eighteenth century, the newly developed steam engine was applied to grain mills. Such advances over the centuries allowed the use of larger and larger stones and, thus, increased the capacity of the mills.

Winnowing (separating the hulls from the groats) was originally accomplished by throwing the mixture into the air on a windy hill, the heavier groats falling onto sheets laid on the ground. Later, this step was done in barns situated so that the doors, open at each end, allowed the prevailing breeze to blow away the hulls (Lockhart 1983). A variety of home kilns were also developed to dry oat grains, rendering them easier to mill and imparting a toasty flavor.

The next major advance in oat milling came with the 1875 invention of a groat-cutting machine by Asmus Ehrlichson in Akron, Ohio. The groats could now be cut into uniform pieces for a higher quality meal. Prior to this development, the crushing of groats had meant a mixture of fine flour and more or

less coarse bits of endosperm that made an inferior meal when cooked. Steel-cut oats were also less liable to become rancid than the crushed grain. Steel-cut oats, available today as Scotch oats, were quite popular until superseded by the innovation of oat flakes.

Rollers, known as far back as the 1650s, were used for crushing groats, much as stones had been used. But in the 1870s it was discovered that when partially cooked groats were rolled, they formed flakes. The production and marketing of oat flakes, which began in the 1880s with a pair of small oat processors, was adopted by the (then fledgling) Quaker Oats Company (Thornton 1933: 149–52). Moreover, steel-cut oats as well as whole groats could be flaked, the former producing a faster-cooking product because of its thinner, smaller flakes.

Stones were used to remove oat hulls up until about 1936, when they were replaced with impact hullers. Impact hulling involves introducing the oats to the center of a spinning rotor that propels them outward against a carborundum or rubber ring. The impact removes the hull with minimum groat breakage. This huller has a better groat yield and is more energy efficient than stones (Deane and Commers 1986).

More recent developments in oat products include instant oats, flaked so thin that they cook by the addition of boiling water, and oat bran. Oat bran, the coarse fraction produced by sieving ground groats, contains a higher proportion of soluble fiber (predominantly beta-glucan), useful for lowering high levels of cholesterol (Ripsin et al. 1992).

Current practice in milling oats has been detailed by D. Deane and E. Commers (1986) and by D. Burnette and colleagues (1992) (Figure II.A.6.1). The first steps involve cleaning and grading. Other grains, foreign matter, and weed seeds are removed by a series of separations according to size and density on screens, disc separators, graders, and aspirators. At the same time, oat is separated into milling grade and other grades (light oats, pin oats, slim oats, and double oats), which are used for animal feed. The milling-grade oat is then subjected to drying in ovens to reduce the moisture from about 13 percent to 6 to 7 percent, followed by cooling. Alternatively, the drying may be delayed until after the hulls are removed. Dried oat has tougher groats and is less subject to breakage during the dehulling process.

The huller produces a mixture of groats, hulls, and broken pieces, and these are separated by air aspiration. The groats are separated by size and passed on to the cutting or flaking machinery. Groats that are steel cut into two to four pieces formerly were available as Scotch oats but are now used mostly for other products. The whole and the steel-cut groats are steamed and rolled, producing regular or quick-cooking flakes, respectively. Oat flour is made by grinding oat flakes, steel-cut groats, or middlings. It is used in ready-to-eat breakfast foods and other products. Oat bran is produced by sieving coarsely ground oat flour.

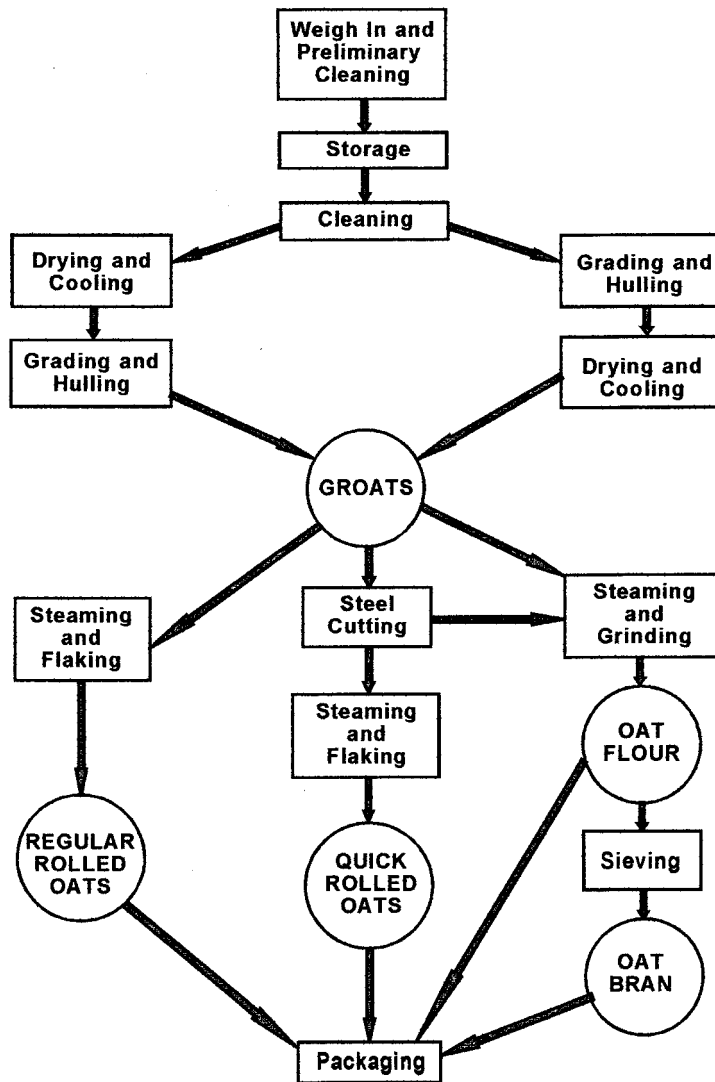


Figure II.A.6.1. Flow diagram of typical oat-milling sequence. (From Deane and Commers 1986.)

Uses of Oat

Although archaeological records indicate that primitive peoples employed oat as a food source, the first written reference to its use was Pliny's observation that the Germans knew oat well and "made their porridge of nothing else" (Rackham 1950). Oatmeal porridge was an acknowledged Scottish staple as early as the fifth century A.D. (Kelly 1975). Porridge was prepared by boiling oatmeal in water, and it was consumed with milk, and sometimes honey, syrup, or treacle (Lockhart 1983). Brose, made by adding boiling water to oatmeal, was of a thicker consistency. In Ireland during the same period, oatmeal porridge was consumed in a mixture with honey and butter or milk (Joyce 1913). Popular in Scotland were oatcakes, prepared by making a

dough of oatmeal and water and heating it on a baking stone or griddle, and, in fact, oatcakes are still produced in Scotland by commercial bakeries as well as in the home. In England, a fourteenth-century tale recounted that in times of economic stress, the poor of London ate a gruel of oatmeal and milk (Langland 1968), and in 1597, J. Gerrard indicated that oat was used to make bread and cakes as well as drink in northeast England (Woodward 1931). Because oat flour lacks gluten and produces a flat cake, it must be mixed with wheat flour for bread-making. This was probably a common practice, which extended the quantity of the more valuable, and perhaps less productive, wheat. Gerrard also described medicinal uses of oat to improve the complexion and as a poultice to cure a "stitch" (Woodward 1931).

Young (1892) noted that potatoes (*Solanum tuberosum* L.) and milk were the staple foods of the common people in most of Ireland, but this diet was supplemented occasionally with oatmeal. Following the potato crop failure of 1740, however, oatmeal became the main ingredient in publicly provided emergency foods (Drake 1968).

Both Young and Adam Smith (1776) discussed the diets of the common people of the time. Smith was critical of the heavy dependence on oat in Scotland and believed that potatoes or wheat bread staples produced a healthier population. Young was not critical of oat, but he believed that the relatively healthy rural population in Ireland resulted from consumption of milk in addition to potatoes, rather than the more commonplace ale or tea consumed in England. In mid-nineteenth-century England, the highest-paid factory workers ate meat daily, whereas the poorest ate cheese, bread, oatmeal porridge, and potatoes (Engels 1844). However, oatmeal was a popular breakfast food among the wealthy.

Although oat was produced in the North American colonies from the time of the earliest settlements, it was not considered a human food except in a few predominantly Scottish settlements. A small quantity of oat was imported from Europe and typically sold in drug stores to invalids and convalescents. That oat had medicinal value had been known since Roman times (Woodward 1931; Font Quer 1962), but it was believed, erroneously, that domestically produced oat was not suitable for human consumption. Most nineteenth-century cookbooks in the United States either contained no recipes for oatmeal or suggested it as food for the infirm (Webster 1986). Indeed, the idea of humans consuming oats was a subject of ridicule by humorists and cartoonists in several national publications (Thornton 1933).

The selling of domestically produced oatmeal for human consumption in the United States began in earnest at the end of the nineteenth century, and its increasing popularity with the public can be attributed to the improved technology of producing rolled oat flakes, selling them in packages instead of in bulk, and a marketing strategy of portraying oatmeal as a healthful and nutritious product (Thornton 1933). The story of the marketing of oatmeal to the North American public is notable because it represented the first use of mass marketing techniques that are commonplace today (Marquette 1967).

New food uses for oat continue to be developed and marketed to a generally receptive public. The popularity of ready-to-eat breakfast cereals, many of which are oat based or contain some oat, has contributed to the increased food demand for oat. In the hot cereal market, instant oat products are achieving a greater market share due to consumers' preference for quick breakfast products requiring little, if any, preparation. Research on the effects of oat bran on blood cholesterol levels has also increased demand

for oat bran products from health-conscious consumers. Oat is a popular ingredient in breads, cookies, and infant foods.

Nutritional Value

The nutritional value of oat has long been recognized. Although there was no scientific basis for nutritional claims in the Middle Ages, surely it was known that a staple diet of oat supported people accustomed to hard physical labor. Jean Froissart, a historian of the fourteenth century, wrote that Scottish soldiers carried with them, on their horses, bags of oat and metal plates upon which to cook oatcakes (Lockhart 1983). Oat consumption increased markedly in Scotland in the eighteenth century, coincident with a drop in meat consumption (Symon 1959), and much of the Scottish diet during the eighteenth century was oat.

As the science of nutrition developed in the twentieth century, scientists began to measure human needs for vitamins, minerals, essential amino acids and fatty acids, and energy. Foods were analyzed to ascertain their content of these essentials, and cereals, in general, and oat, in particular, were recognized as important contributors to human nutrition. But because of certain deficiencies, grains by themselves could not be considered "complete" foods.

The primary constituent of oat is starch, which constitutes from 45 to 62 percent of the groat by weight (Paton 1977). This percentage is lower than that of other cereals because of the higher levels of protein, fiber, and fat. Oat starch is highly digestible, and oat is a good energy source. Oat protein is higher than that of most other cereals (15 to 20 percent, groat basis) and contains a better balance of essential amino acids (Robbins, Pomeranz, and Briggles 1971). Nevertheless, lysine, threonine, and methionine are contained in less than optimal proportions. The oil content of oat is also higher than that of other cereals, ranging from about 5 to 9 percent for cultivated varieties (Youngs 1986), but genotypes with extreme values have been identified (Brown and Craddock 1972; Schipper and Frey 1991b). Oat oil is nutritionally favorable because of a high proportion of unsaturated fatty acids, including the essential fatty acid, linoleic acid.

The mineral content of oat is typical of that of other cereals (Peterson et al. 1975). Oat provides a significant proportion of manganese, magnesium, and iron and is also a source of zinc, calcium, and copper. Although high in phosphorus, much of this is unavailable as phytic acid. Oat also contains significant amounts of several vitamins - thiamin, folic acid, biotin, pantothenic acid, and vitamin E (Lockhart and Hurt 1986) - but contains little or no vitamins A, C, and D.

In developed countries where food for most people is abundant, the emphasis in nutrition has changed from correcting nutrient deficiencies to avoiding

excessive consumption of saturated fats, refined sugar, and cholesterol while consuming foods high in carbohydrate and fiber. Diets containing whole-grain cereals fit well into this prescription for healthful eating. Oat, along with barley, contains a relatively high amount of beta-glucan, a soluble fiber that has been shown in numerous studies to lower the cholesterol levels of hypercholesterolemic subjects (Ripsin et al. 1992).

This knowledge spawned a plethora of products made from oat bran, because it was established that the bran fraction contained a higher concentration of beta-glucan than did whole oat. Although the marketplace has now discarded a number of these products that contained nutritionally insignificant quantities of oat bran, there is a definite place for oat bran in therapy for high blood cholesterol.

David M. Peterson
J. Paul Murphy

Bibliography

- Aung, T., and H. Thomas. 1978. The structure and breeding behavior of a translocation involving the transfer of powdery mildew resistance from *Avena barbata* Pott. into the cultivated oat. *Euphytica* 27: 731-9.
- Baker, R. J., and R. I. H. McKenzie. 1972. Heritability of oil content in oats, *Avena sativa* L. *Crop Science* 12: 201-2.
- Bar-Yosef, O., and M. E. Kislev. 1989. Early farming communities in the Jordan Valley. In *Foraging and farming*, ed. D. R. Harris and G. C. Hillman, 632-42. Winchester, Mass.
- Baum, B. R. 1977. *Oats: Wild and cultivated*. Ottawa.
- Bennett, R., and J. Elton. 1898. *History of corn milling*, Vol. 1. London and Liverpool.
- Bidwell, P. W., and J. I. Falconer. 1925. *History of agriculture in the northern United States, 1620-1860*. Washington, D.C.
- Branson, C. V., and K. J. Frey. 1989. Recurrent selection for groat oil content in oat. *Crop Science* 29: 1382-7.
- Brown, C. M., D. E. Alexander, and S. G. Carmer. 1966. Variation in oil content and its relation to other characters in oats (*Avena sativa* L.). *Crop Science* 6: 190-1.
- Brown, C. M., and J. C. Craddock. 1972. Oil content and groat weight of entries in the world oat collection. *Crop Science* 12: 514-15.
- Brownlee, H. J., and F. L. Gunderson. 1938. Oats and oat products. *Cereal Chemistry* 15: 257-72.
- Burnette, D., M. Lenz, P. F. Sisson, et al. 1992. Marketing, processing, and uses of oat for food. In *Oat science and technology*, ed. H. G. Marshall and M. E. Sorrells, 247-63. Madison, Wis.
- Campbell, A. R., and K. J. Frey. 1972. Inheritance of groat protein in interspecific oat crosses. *Canadian Journal of Plant Science* 52: 735-42.
- Candolle, A. de. 1886. *Origin of cultivated plants*. New York.
- Coffman, F. A. 1946. Origin of cultivated oats. *Journal of the American Society of Agronomy* 38: 983-1002.
1977. *Oat history, identification and classification*. Technical Bulletin No. 1516. Washington, D.C.
- Coffman, F. A., H. C. Murphy, and W. H. Chapman. 1961. Oat breeding. In *Oats and oat improvement*, ed. F. A. Coffman, 263-329. Madison, Wis.
- Cox, T. S., and K. J. Frey. 1985. Complementarity of genes for high groat protein percentage from *Avena sativa* L. and *A. sterilis* L. *Crop Science* 25: 106-9.
- Darby, W. J., P. Ghalioungui, and L. Grivetti. 1977. *Food: The gift of Osiris*. New York.
- Deane, D., and E. Commers. 1986. Oat cleaning and processing. In *Oats: Chemistry and technology*, ed. F. H. Webster, 371-412. St. Paul, Minn.
- Drake, M. 1968. The Irish demographic crisis of 1740-41. In *Historical studies*, ed. T. W. Moody, 101-24. New York.
- Ellsworth, H. W. 1838. *Valley of the upper Wabash, Indiana*. New York.
- Engels, F. [1844] 1968. *The condition of the working class in England*. Stanford, Calif.
- Font Quer, P. 1962. *Plantas medicinales*. Madrid.
- Forsberg, R. A., V. L. Youngs, and H. L. Shands. 1974. Correlations among chemical and agronomic characteristics in certain oat cultivars and selections. *Crop Science* 14: 221-4.
- Frey, K. J., J. A. Browning, and M. D. Simons. 1985. Registration of 'Multiline E76' and 'Multiline E77' oats. *Crop Science* 25: 1125.
- Frey, K. J., and E. G. Hammond. 1975. Genetics, characteristics, and utilization of oil in caryopses of oat species. *Journal of the American Oil Chemists' Society* 52: 358-62.
- Frey, K. J., E. G. Hammond, and P. K. Lawrence. 1975. Inheritance of oil percentage in interspecific crosses of hexaploid oats. *Crop Science* 15: 94-5.
- Grey, L. C., and E. K. Thompson. 1933. *History of agriculture in the southern United States to 1860*. Washington, D.C.
- Hansen, J. R., and J. M. Renfrew. 1978. Palaeolithic-Neolithic seed remains at Franchthi Cave, Greece. *Nature* (London) 271: 349-52.
- Harlan, J. R. 1977. The origins of cereal agriculture in the old world. In *Origins of agriculture*, ed. C. A. Reed, 357-83. The Hague.
1989. Wild-grass seed harvesting in the Sahara and sub-Saharan of Africa. In *Foraging and farming*, ed. D. R. Harris and G. C. Hillman, 79-98. Winchester, Mass.
- Harvey, P. D. A. 1965. *A medieval Oxfordshire village: Cuxham 1240 to 1400*. London.
- Hillman, G. C., S. M. Colledge, and D. R. Harris. 1989. Plant-food economy during the Epipalaeolithic period at Tell Abu Hureyra, Syria: Dietary diversity, seasonality and modes of exploitation. In *Foraging and farming*, ed. D. R. Harris and G. C. Hillman, 240-68. Winchester, Mass.
- Holden, T. G. 1986. Preliminary report of the detailed analyses of the macroscopic remains from the gut of Lindow man. In *Lindow man, the body in the bog*, ed. I. M. Stead, J. B. Bourke, and D. Brothwell, 116-25. Ithaca, N.Y.
- Hopf, M. 1969. Plant remains and early farming in Jericho. In *The domestication and exploitation of plants and animals*, ed. P. J. Ucko and G. W. Dimbleby, 355-9. Chicago.
- Hutchinson, J. B., and H. F. Martin. 1955. The chemical composition of oats. I. The oil and free fatty acid content of oats and groats. *Journal of Agricultural Science* 45: 411-18.
- Iwig, M. M., and H. W. Ohm. 1976. Genetic control of protein from *Avena sterilis* L. *Crop Science* 16: 749-52.

- Joyce, P. W. 1913. *A social history of ancient Ireland*, Vol. 2. New York.
- Karow, R. S., and R. A. Forsberg. 1984. Oil composition in parental, F_1 and F_2 populations of two oat crosses. *Crop Science* 24: 629-32.
1985. Selection for linoleic acid concentration among progeny of a high $\times$ low linoleic acid oat cross. *Crop Science* 15: 45-7.
- Kelly, J. N. D. 1975. *Jerome*. New York.
- Kuenzel, K. A., and K. J. Frey. 1985. Protein yield of oats as determined by protein percentage and grain yield. *Euphytica* 34: 21-31.
- Ladizinsky, G. 1969. New evidence on the origin of the hexaploid oats. *Evolution* 23: 676-84.
1988. The domestication and history of oats. In *Proceedings of the Third International Oat Conference* (Lund, Sweden, July 4-8, 1988), ed. B. Mattsson and R. Lyhagen, 7-12. Svalöf, Sweden.
- Langland, W. 1968. *The vision of Piers ploughman*. New York.
- Lawes, D. A. 1977. Yield improvement in spring oats. *Journal of Agricultural Science* 89: 751-7.
- Leggett, J. M. 1992. Classification and speciation in *Avena*. In *Oat science and technology*, ed. H. G. Marshall and M. E. Sorrells, 29-52. Madison, Wis.
- Lockhart, G. W. 1983. *The Scot and his oats*. Barr, Ayrshire, U.K.
- Lockhart, H. B., and H. D. Hurt. 1986. Nutrition of oats. In *Oats chemistry and technology*, ed. Francis H. Webster, 297-308. St. Paul, Minn.
- Luby, J. J., and D. D. Stuthman. 1983. Evaluation of *Avena sativa* L./*A. fatua* L. progenies for agronomic and grain quality characters. *Crop Science* 23: 1047-52.
- Lynch, P. J., and K. J. Frey. 1993. Genetic improvement in agronomic and physiological traits of oat since 1914. *Crop Science* 33: 984-8.
- Marquette, A. F. 1967. *Brands, trademarks and goodwill: The story of the Quaker Oats Company*. New York.
- Marshall, H. G. 1992. Breeding oat for resistance to environmental stress. In *Oat science and technology*, ed. H. G. Marshall and M. E. Sorrells, 699-749. Madison, Wis.
- McFerson, J. K., and K. J. Frey. 1991. Recurrent selection for protein yield in oat. *Crop Science* 31: 1-8.
- McMullen, M. S., and F. L. Patterson. 1992. Oat cultivar development in the USA and Canada. In *Oat science and technology*, ed. H. G. Marshall and M. E. Sorrells, 573-612. Madison, Wis.
- Moser, H. S., and K. J. Frey. 1994. Yield component responses associated with increased groat yield after recurrent selection in oat. *Crop Science* 34: 915-22.
- Murphy, J. P., and L. A. Hoffman. 1992. The origin, history, and production of oat. In *Oat science and technology*, ed. H. G. Marshall and M. E. Sorrells, 1-28. Madison, Wis.
- Paton, D. 1977. Oat starch. Part 1. Extraction, purification and pasting properties. *Stärke* 29: 149-53.
- Peterson, D. M. 1976. Protein concentration, concentration of protein fractions and amino acid balance in oats. *Crop Science* 16: 663-6.
1991. Genotype and environmental effects on oat beta-glucan concentration. *Crop Science* 31: 1517-20.
1992. Composition and nutritional characteristics of oat grain and products. In *Oat science and technology*, ed. H. G. Marshall and M. E. Sorrells, 265-92. Madison, Wis.
- Peterson, D. M., J. Senturia, V. L. Youngs, and L. E. Schrader. 1975. Elemental composition of oat groats. *Journal of Agricultural and Food Chemistry* 23: 9-13.
- Peterson, D. M., D. M. Wesenberg, and D. E. Burrup. 1995. β -glucan content and its relationship to agronomic characteristics in elite oat germplasm. *Crop Science* 35: 965-70.
- Rackham, H. 1950. *Pliny natural history*, Vol. 5. Books 17-19. Cambridge, Mass.
- Reich, J. M., and M. A. Brinkman. 1984. Inheritance of groat protein percentage in *Avena sativa* L. $\times$ *A. fatua* L. crosses. *Euphytica* 33: 907-13.
- Renfrew, J. M. 1969. The archaeological evidence for the domestication of plants: Methods and problems. In *The domestication and exploitation of plants and animals*, ed. P. J. Ucko and G. W. Dimbleby, 149-72. Chicago.
- Reysack, J. J., D. D. Stuthman, and R. E. Stucker. 1993. Recurrent selection in oat: Stability of yield and changes in unselected traits. *Crop Science* 33: 919-24.
- Ripsin, C. M., J. M. Keenan, D. R. Jacobs, Jr., et al. 1992. Oat products and lipid lowering: A meta-analysis. *Journal of the American Medical Association* 267: 3317-25.
- Robbins, G. S., Y. Pomeranz, and L. W. Briggles. 1971. Amino acid composition of oat groats. *Journal of Agricultural and Food Chemistry* 19: 536-9.
- Roche, I. A. de la, V. D. Burrows, and R. I. H. McKenzie. 1977. Variation in lipid composition among strains of oats. *Crop Science* 17: 145-8.
- Rodgers, D. M., J. P. Murphy, and K. J. Frey. 1983. Impact of plant breeding on the grain yield and genetic diversity of spring oats. *Crop Science* 23: 737-40.
- Schipper, H., and K. J. Frey. 1991a. Selection for groat-oil content in oat grown in field and greenhouse. *Crop Science* 31: 661-5.
- 1991b. Observed gains from three recurrent selection regimes for increased groat-oil content of oat. *Crop Science* 31: 1505-10.
- Schrackel, D. J., V. D. Burrows, and J. A. Ingemansen. 1992. Harvesting, storing, and feeding of oat. In *Oat science and technology*, ed. H. G. Marshall and M. E. Sorrells, 223-45. Madison, Wis.
- Sharma, D. C., and R. A. Forsberg. 1977. Spontaneous and induced interspecific gene transfer for crown rust resistance in oats. *Crop Science* 17: 855-60.
- Smith, A. 1776. *The wealth of nations*, Vol. 1. London.
- Sorrells, M. E., and S. R. Simmons. 1992. Influence of environment on the development and adaptation of oat. In *Oat science and technology*, ed. H. G. Marshall and M. E. Sorrells, 115-63. Madison, Wis.
- Stanton, T. R. 1936. Superior germ plasm in oats. In *USDA yearbook of agriculture*, 347-414. Washington, D.C.
- Sturtevant, E. L. 1919. *Sturtevant's notes on edible plants*, ed. U. P. Hendrick. Albany, N.Y.
- Symon, J. A. 1959. *Scottish farming*. Edinburgh.
- Takeda, K., and K. J. Frey. 1976. Contributions of vegetative growth rate and harvest index to grain yield of progenies from *Avena sativa* $\times$ *A. sterilis* crosses. *Crop Science* 16: 817-21.
- Thomas, H., J. M. Haki, and S. Arangzeb. 1980. The introgression of characters of the wild oat *Avena magna* ($2n = 4x = 28$) into cultivated oat *A. sativa* ($2n = 6x = 42$). *Euphytica* 29: 391-9.
- Thornton, H. J. 1933. *The history of The Quaker Oats Company*. Chicago.
- Thro, A. M., and K. J. Frey. 1985. Inheritance of groat oil content and high oil selections in oats (*Avena sativa* L.). *Euphytica* 34: 251-63.
- Thro, A. M., K. J. Frey, and E. G. Hammond. 1983. Inheritance of fatty acid composition in oat (*Avena sativa* L.). *Qualitas Plantarum, Plant Foods for Human Nutrition* 32: 29-36.

- Villaret-von Rochow, M. 1971. *Avena ludoviciana* Dur. im Schweizer Spätneolithikum, ein Beitrag zur Abstammung des Saathafers. *Berichte der Deutschen Botanischen Gesellschaft* 84: 243-8.
- Weaver, S. H. 1988. The history of oat milling. In *Proceedings of the Third International Oat Conference* (Lund, Sweden, July 4-8, 1988), ed. B. Mattsson and R. Lyhagen, 47-50. Svalöf, Sweden.
- Webster, F. H. 1986. Oat utilization: Past, present, and future. In *Oats: Chemistry and technology*, ed. F. H. Webster, 413-26. St. Paul, Minn.
- White, K. D. 1970. *Roman farming*. Ithaca, N.Y.
- Woodward, M. 1931. *Leaves from Gerrard's herbal*. New York.
- Wych, R. D., and D. D. Stuthman. 1983. Genetic improvement in Minnesota - Adapted oat cultivars released since 1923. *Crop Science* 23: 879-81.
- Yanushevich, Z. V. 1989. Agricultural evolution north of the Black Sea from Neolithic to the Iron Age. In *Foraging and farming*, ed. D. R. Harris and G. C. Hillman, 606-19. Winchester, Mass.
- Young, A. 1892. *Arthur Young's tour in Ireland (1776-1779)*. London.
- Youngs, V. L. 1986. Oat lipids and lipid-related enzymes. In *Oats: Chemistry and technology*, ed. F. H. Webster, 205-26. St. Paul, Minn.
- Youngs, V. L., and M. Püskülcü. 1976. Variation in fatty acid composition of oat groats from different cultivars. *Crop Science* 16: 881-3.
- Youngs, V. L., M. Püskülcü, and R. R. Smith. 1977. Oat lipids. 1. Composition and distribution of lipid components in two oat cultivars. *Cereal Chemistry* 54: 803-12.
- Zohary, D., and M. Hopf. 1988. *Domestication of plants in the old world*. Oxford.
- Zohary, M. 1982. *Plants of the Bible*. New York.

II.A.7 🌾 Rice

Economic and Biological Importance of Rice

Rice in Human Life

Among the cereals, rice and wheat share equal importance as leading food sources for humankind. Rice is a staple food for nearly one-half of the world's population. In 1990, the crop was grown on 145.8 million hectares of land, and production amounted to 518.8 million metric tons of grain (paddy, rough rice). Although rice is grown in 112 countries, spanning an area from 53° latitude north to 35° south, about 95 percent of the crop is grown and consumed in Asia. Rice provides fully 60 percent of the food intake in Southeast Asia and about 35 percent in East Asia and South Asia. The highest level of per capita rice consumption (130 to 180 kilograms [kg] per year, 55 to 80 percent of total caloric source) takes place in Bangladesh, Cambodia, Indonesia, Laos, Myanmar (Burma), Thailand, and Vietnam.

Although rice commands a higher price than wheat on the international market, less than five per-

cent of the world's rice enters that market, contrasted with about 16 percent of the wheat. Low-income countries, China and Pakistan, for example, often import wheat at a cheaper price and export their rice.

Biological Value in Human Nutrition

Although rice has a relatively low protein content (about 8 percent in brown rice and 7 percent in milled rice versus 10 percent in wheat), brown rice (caryopsis) ranks higher than wheat in available carbohydrates, digestible energy (kilojoules [kJ] per 100 grams), and net protein utilization. Rice protein is superior in lysine content to wheat, corn, and sorghum. Milled rice has a lower crude fiber content than any other cereal, making rice powder in the boiled form suitable as infant food. For laboring adults, milled rice alone could meet the daily carbohydrate and protein needs for sustenance although it is low in riboflavin and thiamine content. For growing children, rice needs to be supplemented by other protein sources (Hegsted 1969; Juliano 1985b).

The Growing Importance of Rice

On the basis of mean grain yield, rice crops produce more food energy and protein supply per hectare than wheat and maize. Hence, rice can support more people per unit of land than the two other staples (Lu and Chang 1980). It is, therefore, not surprising to find a close relationship in human history between an expansion in rice cultivation and a rapid rise in population growth (Chang 1987).

As a human food, rice continues to gain popularity in many parts of the world where other coarse cereals, such as maize, sorghum and millet, or tubers and roots like potatoes, yams, and cassava have traditionally dominated. For example, of all the world's regions, Africa has had the sharpest rise in rice consumption during the last few decades.

Rice for table use is easy to prepare. Its soft texture pleases the palate and the stomach. The ranking order of food preference in Asia is rice,

followed by wheat, maize, and the sweet potato; in Africa it is rice or wheat, followed by maize, yams, and cassava (author's personal observation).

In industrial usage, rice is also gaining importance in the making of infant foods, snack foods, breakfast cereals, beer, fermented products, and rice bran oil, and rice wine remains a major alcoholic beverage in East Asia. The coarse and silica-rich rice hull is finding new use in construction



materials. Rice straw is used less in rope and paper making than before, but except for modern varieties, it still serves as an important cattle feed throughout Asia. Because rice flour is nearly pure starch and free from allergens, it is the main component of face powders and infant formulas. Its low fiber content has led to an increased use of rice powder in polishing camera lenses and expensive jewelry.

Botany, Origin, and Evolution

Botany

Rice is a member of the grass family (Gramineae) and belongs to the genus *Oryza* under tribe Oryzeae. The genus *Oryza* includes 20 wild species and 2 cultivated species (cultigens). The wild species are widely distributed in the humid tropics and subtropics of Africa, Asia, Central and South America, and Australia (Chang 1985). Of the two cultivated species, African rice (*O. glaberrima* Steud.) is confined to West Africa, whereas common or Asian rice (*O. sativa* L.) is now commercially grown in 112 countries, covering all continents (Bertin et al. 1971).

The wild species have both diploid ($2n = 2x = 24$) and tetraploid ($2n = 4x = 48$) forms, while the two cultigens are diploid and share a common genome (chromosome group). Incompatibility exists among species having different genomes. Partial sterility also shows up in hybrids when different ecogeographic races of *O. sativa* are hybridized. The cultivated species of *Oryza* may be classified as semiaquatic plants, although extreme variants are grown not only in deep water (up to 5 meters) but also on dry land (Chang 1985).

Among the cereals, rice has the lowest water use efficiency. Therefore, rice cannot compete with dryland cereals in areas of low rainfall unless irrigation water is readily available from reservoirs, bunds, and the like. On the other hand, the highest yields of traditional varieties have been obtained in regions of cloudless skies, such as in Spain, California, and northern Japan (Lu and Chang 1980).

The "wild rice" of North America is *Zizania palustris* (formerly *Z. aquatica* L. [$2n = 30$]), which belongs to one of the 11 related genera in the same tribe. Traditionally, this species was self-propagating and harvested only by Native Americans in the Great Lakes area. Now it is commercially grown in Minnesota and northern California.

Origin

The origin of rice was long shrouded by disparate postulates because of the pantropical but disjunct distribution of the 20 wild species across four continents, the variations in characterizing and naming plant specimens, and the traditional feud concerning the relative antiquity of rice in India versus China. Among the botanists, R. J. Roschevitz (1931) first postulated that the center of origin of the section *Sativa*

Roschev., to which *O. glaberrima* and *O. sativa* belong, was in Africa and that *O. sativa* had originated from multiple species. A divergent array of wild species was proposed by different workers as the putative ancestor of *O. sativa* (Chang 1976b).

Several workers considered "*O. perennis* Moench" (an ambiguous designation of varying applications) as the common progenitor of both cultigens (Chang 1976b). A large number of scholars had argued that Asian rice originated in the Indian subcontinent (South Asia), although A. de Candolle (1884), while conceding that India was more likely the original home, considered China to have had an earlier history of rice cultivation.

On the basis of historical records and the existence of wild rices in China, Chinese scholars maintained that rice cultivation was practiced in north China during the mythological Sheng Nung period (c. 2700 B.C.) and that *O. sativa* of China evolved from wild rices (Ting 1961). The finding of rice glume imprints at Yang-shao site in north China (c. 3200–2500 B.C.) during the 1920s reinforced the popular belief that China was one of the centers of its origin (Chinese Academy of Agricultural Sciences 1986).

Since the 1950s, however, rice researchers have generally agreed that each of the two cultigens originated from a single wild species. But disputes concerning the immediate ancestor of *O. sativa* persist to this day (Chang 1976b, 1985; Oka 1988). A multidisciplinary analysis of the geographic distribution of the wild species and their genomic composition in relation to the "Glossopterid Line" (northern boundary) of the Gondwanaland fragments (Melville 1966) strongly indicated the Gondwanaland origin of the genus *Oryza* (Chang 1976a, 1976b, 1985). This postulate of rice having a common progenitor in the humid zones of the supercontinent Pangaea before it fractured and drifted apart can also explain the parallel evolutionary pattern of the two cultigens in Africa and Asia respectively. It also reconciles the presence of closely related wild species having the same genome in Australia and in Central and South America. Thus, the antiquity of the genus dates back to the early Cretaceous period of more than 130 million years ago.

Evolution

The parallel evolutionary pathway of *O. glaberrima* in Africa and of *O. sativa* in Asia was from perennial wild → annual wild → annual cultigen, a pattern common to other grasses and many crop plants. The parallel pathways are:

Africa: *O. longistaminata* → *O. barthii* → *O. glaberrima*.

Asia: *O. rufipogon* → *O. nivara* → *O. sativa*.

This scheme can resolve much that has characterized past disputes on the putative ancestors of the

two cultigens. Wild perennial and annual forms having the same A genome are present in Australia and in Central and South America, but the lack of incipient agriculture in Australia and of wetland agronomy in tropical America in prehistoric times disrupted the final step in producing an annual cultigen.

It needs to be pointed out that the putative ancestors, especially those in tropical Asia, are conceptually wild forms of the distant past, because centuries of habitat disturbance, natural hybridization, and dispersal by humans have altered the genetic structure of the truly wild ancestors. Most of the wild rices found in nature today are hybrid derivatives of various kinds (Chang 1976b; 1985). The continuous arrays of variants in natural populations have impaired definitive studies on the wild progenies (Chang 1976b; Oka 1988).

The differentiation and diversification of annual wild forms into the early prototypes of cultigen in South and mainland Southeast Asia were accelerated by marked climatic changes during the Neothermal age of about 10,000 to 15,000 years ago. Initial selection and cultivation could have occurred independently and nearly concurrently at numerous sites within or bordering a broad belt of primary genetic diversity that extends from the Ganges plains below the eastern foothills of Himalaya, through upper Burma, northern Thailand, Laos, and northern Vietnam, to southwest and southern China.

From this belt, geographic dispersal by various agents, particularly water currents and humans, lent impetus to ecogenetic differentiation and diversification under human cultivation. In areas inside China where winter temperatures fell below freezing, the cultivated forms (cultivars) became true domesticates, depending entirely on human care for their perpetuation and propagation. In a parallel manner, the water buffalo was brought from the swamps of the south into the northern areas and coevolved as another domesticate (Chang 1976a).

In West Africa, *O. glaberrima* was domesticated from the wild annual *O. barthii* (Chevalier 1932); the latter was adapted primarily to water holes in the savanna and secondarily to the forest zone (Harlan 1973). The cultigen has its most important center of diversity in the central Niger delta. Two secondary centers existed near the Guinean coast (Porteres 1956).

Cultivation of the wild prototypes preceded domestication. Rice grains were initially gathered and consumed by prehistoric people of the humid regions where the perennial plants grew on poorly drained sites. These people also hunted, fished, and gathered other edible plant parts as food. Eventually, however, they developed a liking for the easily cooked and tasty rice and searched for plants that bore larger panicles and heavier grains.

The gathering-and-selection process was more imperative for peoples who lived in areas where sea-

sonal variations in temperature and rainfall were more marked. The earlier maturing rices, which also tend to be drought escaping, would have been selected to suit the increasingly arid weather of the belt of primary diversity during the Neothermal period. By contrast, the more primitive rices of longer maturation, and those, thus, more adapted to vegetative propagation, would have survived better in the humid regions to the south (Chang 1976b; 1985). In some areas of tropical Asia, such as the Jeyapore tract of Orissa State (India), the Batticoloa district (Sri Lanka), and the forested areas of north Thailand, the gathering of free-shattering grains from wild rice can still be witnessed today (Chang 1976b; Higham 1989).

Antiquity of Rice Cultivation

Although the differentiation of the progenitors of *Oryza* species dates back to the early Cretaceous period, the beginning of rice cultivation was viewed by Western scholars as a relatively recent event until extensive excavations were made after the 1950s in China and to a lesser extent in India. Earlier, R. J. Roschevicz (1931) estimated 2800 B.C. as the beginning of rice cultivation in China, whereas the dawn of agriculture in India was attributed to the Harappan civilization, which began about 2500 B.C. (Hutchinson 1976).

Thus far, the oldest evidence from India comes from Koldihwa, U.P., where rice grains were embedded in earthen potsherds and rice husks discovered in ancient cow dung. The age of the Chalcolithic levels was estimated between 6570 and 4530 B.C. (Vishnu-Mittre 1976; Sharma et al. 1980), but the actual age of the rice remains may be as recent as 1500 B.C. (Chang 1987). Another old grain sample came from Mohenjodaro of Pakistan and dates from about 2500 B.C. (Andrus and Mohammed 1958). Rice cultivation probably began in the upper and middle Ganges between 2000 and 1500 B.C. (Candolle 1884; Watabe 1973). It expanded quickly after irrigation works spread from Orissa State to the adjoining areas of Andhra Pradesh and Tamil Nadu in the Iron Age around 300 B.C. (Randhawa 1980).

In Southeast Asia, recent excavations have yielded a number of rice remains dating from 3500 B.C. at Ban Chiang (Thailand); 1400 B.C. at Solana (Philippines); and A.D. 500 at Ban Na Di (Thailand) and at Ulu Leang (Indonesia). Dates between 4000 and 2000 B.C. have been reported from North Vietnam (Dao 1985) but have not yet been authenticated.

These various reports have been summarized by T.T. Chang (1988, 1989a). The widely scattered findings are insufficient to provide a coherent picture of agricultural development in the region, but rice cultivation in mainland Southeast Asia undoubtedly preceded that in insular Southeast Asia (Chang 1988). The paucity of rice-related remains that were confined to

upland sites in northern Thailand could be attributed to the sharp rise in sea level around the Gulf of Thailand during the four millennia between 8000 and 4000 B.C. Floods inundated vast tracts of low-lying land amid which rice chaffs and shell knives for cutting rice stalks were recently found at Khok Phanom Di near the Gulf and dated from 6000 to 4000 B.C. (Higham 1989).

For the Southeast Asian region, several geographers and ethnobotanists had earlier postulated that the cultivation of root crops predated rice culture (Sauer 1952; Spencer 1963; Yen 1977). Yet, this hypothesis falters in view of the apparently rather recent domestication (c. 2000 B.C.) of yams in the region (Alexander and Coursey 1969). In many hilly regions, vegetation probably preceded dryland rice cultivation, but not in wetland areas. In the cooler regions, rice grains were crucial to early cultivators who could store and consume the harvest during the winter months.

Prior to the 1950s, the belief in the antiquity of rice cultivation in China was based on mythical writings in which "Emperor Shen Nung" (c. 2700 B.C.) was supposed to have taught his people to plant five cereals, with rice among them (Candolle 1884; Roschevitz 1931; Ting 1949; Chatterjee 1951). This view, however, was questioned by many non-Chinese botanists and historians because of the paucity of wild rices in China (or rather the paucity of information on the wild rices) and the semiarid environment in north China (Chang 1979b, 1983). Yet in the 1920s, the discovery of rice glume imprints on broken pottery at the Yang-shao site in Henan (Honan) by J. G. Andersson and co-workers (Andersson 1934) was important in linking Chinese archaeology with agriculture. The excavated materials were considered Neolithic in origin and the precise age was not available, though K. C. Chang later gave this author an estimated age of between 3200 and 2500 B.C.

Extensive diggings in the Yangtze basin after the 1950s yielded many rice remains that pushed back rice culture in China even further into antiquity (Chang 1983). The most exciting event was the finding in 1973-4 of carbonized rice kernels, rice straw, bone spades, hoe blades (*ssu*), and cooking utensils that demonstrated a well-developed culture supported by rice cultivation at the He-mu-du (Ho-mu-tu) site in Zhejiang (Chekiang) Province dated at 5005 B.C. (Chekiang Provincial Cultural Management Commission and Chekiang Provincial Museum 1976; Hsia 1977).

The grains were mostly of the *hsien* (*Indica*) type but included some *keng* (*Sinica* or *Japonica*) and intermediate kernels. The discovery also indicated the existence of an advanced rice-based culture in east China that vied in antiquity and sophistication with the millet-based culture in north China as represented by the Pan-po site in Shenxi (Shensi). Another site at Luo-jia-jiao in Zhejiang Province also yielded car-

bonized rice of both ecogeographic races of a similar age estimated at 7000 B.P. (Chang 1989a). In a 1988 excavation at Peng-tou-shan site in Hunan Province, abundant rice husks on pottery or red burnt clay as well as skeletal remains of water buffalo were found. The pottery was dated at between 7150 and 6250 B.C. (uncorrected carbon dating). Diggings in neighboring Hubei (Hupei) Province yielded artifacts of similar age, but the grain type could not be ascertained (Pei 1989). Excavations in Shenxi also produced rice glume imprints on red burnt clay dated between 6000 and 5000 B.C. (Yan 1989).

In contrast to all this scholarly effort on the antiquity of rice cultivation in Asia, our understanding of the matter in West Africa rests solely on the writing of R. Porteres (1956), who dates it from 1500 B.C. in the primary Niger center, and from A.D. 1000 to A.D. 1200 in the two Guinean secondary centers.

Chinese history also recorded that rice culture was well established in Honan and Shenxi Provinces of north China during the Chou Dynasty (1122 to 255 B.C.) by Lungshanoid farmers (Ho 1956; Chang 1968). During the Eastern Chou Dynasty (255 to 249 B.C.), rice was already the staple food crop in the middle and lower basins of the Yangtze River (Ting 1961). Wild rices were amply recorded in historical accounts; their northern limit of distribution reached 38° north latitude (Chang 1983).

Based on the above developments, it appears plausible to place the beginning of rice cultivation in India, China, and other tropical Asian countries at nearly 10,000 years ago or even earlier. Since rice was already cultivated in central and east China at 6000 to 5000 B.C., it would have taken a few millennia for rice to move in from the belt to the south of these regions. The missing links in the history of rice culture in China can be attributed to the dearth of archaeological findings from south China and the relatively recent age of rice remains in southwest China (1820 B.C. at Bei-yan in Yunnan) and south China (2000 B.C. at Shih Hsia in Kwangtung). These areas represent important regions of ecogenetic differentiation or routes of dispersal (Chang 1983).

Linguistic Evidence

A number of scholars have attempted to use etymology as a tool in tracing the origin and dispersal of rice in Asia. The Chinese word for rice in the north, *tao* or *dao* or *dau*, finds its variants in south China and Indochina as *k'au* (for grain), *bao*, *ho*, *beu*, *deu*, and *khaw* (Ting 1961; Chinese Academy of Agricultural Sciences 1986). Indian scholars claimed that the word for rice in Western languages had a Dravidian root and that *ris*, *riz*, *arroz*, rice, *oruza*, and *arrazz* all came from *arisi* (Pankar and Gowda 1976). In insular Southeast Asia, the Austronesian terms *padi* and *paray* for rice and *bras* or *beras* for milled rice predominate (Chinese Academy of Agricultural Sciences 1986; Revel 1988).

On the other hand, Japanese scholars have also emphasized the spread of the Chinese words *ni* or *ne* (for wild rice) and *nu* (for glutinous rice) to Southeast Asia (Yanagita et al. 1969). N. Revel and co-workers (1988) have provided a comprehensive compilation of terms related to the rice plant and its parts derived from the linguistic data of China, Indochina, insular Southeast Asia, and Madagascar. Yet among the different disciplinary approaches, linguistic analyses have not been particularly effective in revealing facts about the dispersal of rice by humans. In part, this is because the ethnological aspects of human migration in the Southeast Asian region remain in a state of flux. (For various viewpoints see *Asian Perspectives* 1988: 26, no.1.)

Geographic Dispersal and Ecogenetic Diversification

Early Dispersal

The early dissemination of rice seeds (grains) could have involved a variety of agents: flowing water, wind, large animals, birds, and humans. The latter have undoubtedly been most effective in directed dispersal: Humans carried rice grains from one place to another as food, seed, merchandise, and gifts. The continuous and varied movements of peoples in Asia since prehistoric times have led to a broad distribution of early *O. sativa* forms, which proliferated in ecogenetic diversification after undergoing the mutation-hybridization-recombination-differentiation cycles and being subjected to both natural and human selection forces at the new sites of cultivation. In contrast, *O. glaberrima* cultivars exhibit markedly less diversity than their Asian counterparts, owing to a shorter history of cultivation and narrower dispersal. The contrast is amplified by other factors as shown in Table II.A.7.1.

Initial dispersal of *O. sativa* from numerous sites in its primary belt of diversity involved a combination of early forms of cultivars and associated wild relatives, often grown in a mixture. Biological findings and historical records point to five generalized routes from the Assam-Meghalaya-Burma region. Rice moved: (1) southward to the southern Bengal Bay area and the southern states of India and eventually

to Sri Lanka; (2) westward to Pakistan and the west coast of India; (3) eastward to mainland Southeast Asia (Indochina); (4) southeastward to Malaysia and the Indonesian islands; and (5) northeastward to southwest China, mainly the Yunnan-Kweichow area, and further into east, central, and south China. The early routes of travel most likely followed the major rivers, namely, Brahmaputra, Ganges, Indus, Mekong, and Yangtze. Routes of sea travel, which came later, were from Thailand and Vietnam to the southern coastal areas of China, from Indonesia to the Philippines and Taiwan, and from China to Japan, as well as from China to Korea to Japan. These routes are summarized in Map II.A.7.1.

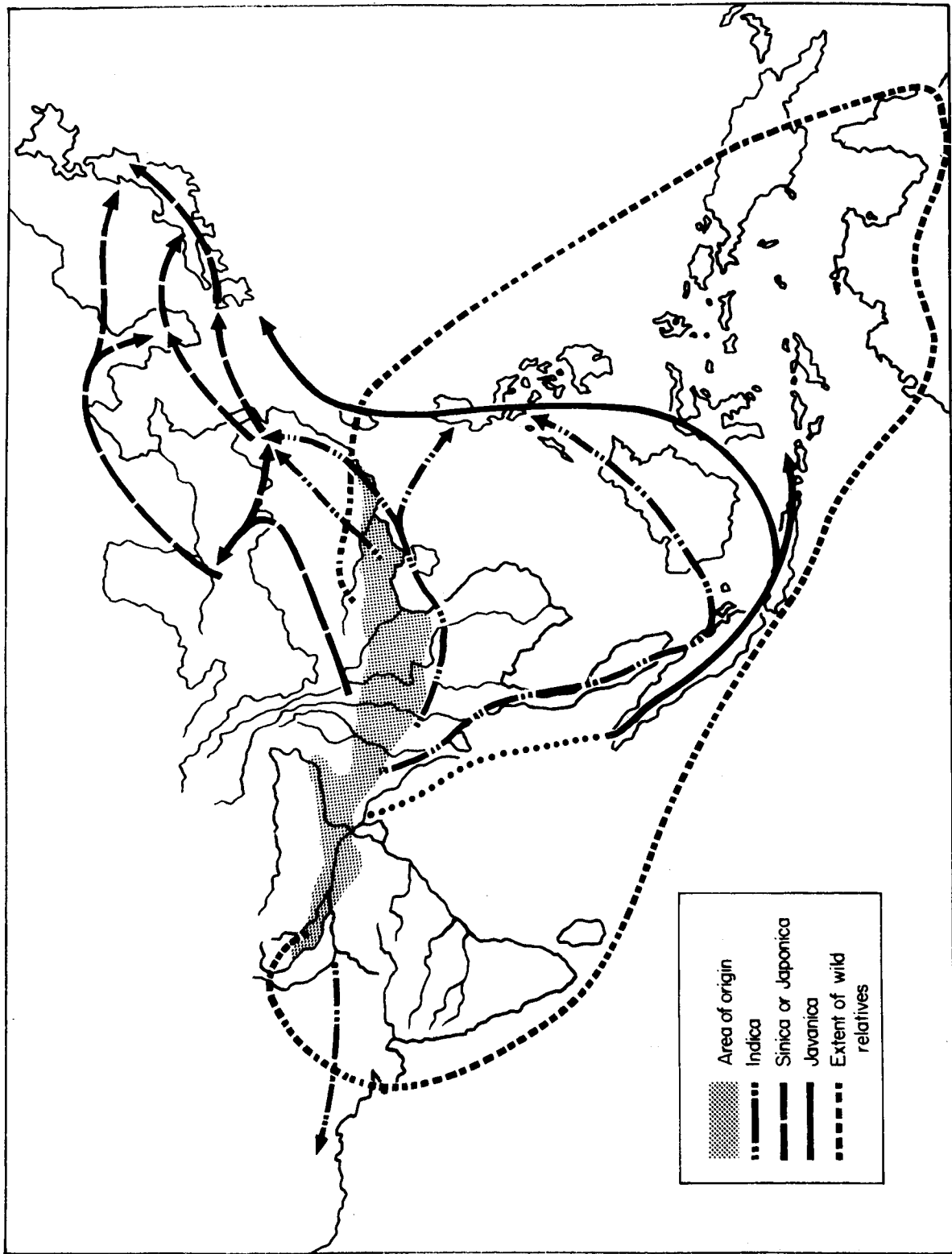
On the basis of ancient samples of rice hulls collected from India and Indochina, covering a span of 10 centuries up to around A.D. 1500, three main groups of cultivars (the Brahmaputra-Gangetic strain, the Bengal strain, and the Mekong strain) have been proposed by T. Watabe (1985). The Mekong strain originating in Yunnan was postulated to have given rise to the Indochina series and the Yangtze River series of cultivars; the latter consisted mainly of the *keng* rices of China. It should be pointed out, however, that the ecogenetic diversification processes following dispersal and the cultivators' preferences could have added complications to the varietal distribution pattern of the present, as later discussions will reveal.

Ecogenetic Differentiation and Diversification

During the early phase of human cultivation and selection, a number of morphological and physiological changes began to emerge. Selection for taller and larger plants resulted in larger leaves, longer and thicker stems, and longer panicles. Subsequent selection for more productive plants and for ease in growing and harvesting led to larger grains. It also resulted in increases in: (1) the rate of seedling growth; (2) tillering capacity; (3) the number of leaves per tiller and the rate of leaf development; (4) the synchronization of tiller development and panicle formation (for uniform maturation); (5) the number of secondary branches on a panicle; and (6) panicle weight (a product of spikelet number and grain weight). Concurrently, there were decreases or losses of the primitive features, such as: (1) rhizome formation; (2) pigmentation of plant parts; (3) awn length; (4) shattering of grains from the panicle; (5) growth duration; (6) intensity of grain dormancy; (7) response to short day length; (8) sensitivity to low temperatures; and (9) ability to survive in flood waters. The frequency of cross pollination also decreased so that the plants became more inbred and increasingly dependent on the cultivators for their propagation (by seed) and perpetuation (by short planting cycles) (Chang 1976b).

Table II.A.7.1 Contrast in diversification: *Oryza sativa* vs. *glaberrima*

Factor	Asia	W. Africa
Latitudinal spread	10° C-53° N	5° N-17° N
Topography	Hilly	Flat
Population density	High	Low
Movement of people	Continuous	Little
Iron tools	Many	None or few
Draft animals	Water buffalo and oxen	?



Map IIA.7.1. Extent of wild relatives and spread of *O. sativa* in Asia and Oceania. (Adapted from Chang 1976b.)

When rice cultivars were carried up and down along the latitudinal or altitudinal clines or both, the enormous genetic variability in the plants was released, and the resulting variants expressed their new genetic makeup while reacting to changing environmental factors. The major environmental forces are soil properties, water supply, solar radiation intensity, day length, and temperature range, especially the minimum night temperatures. Those plants that could thrive or survive in a new environment would become fixed to form an adapted population – the beginning of a new ecostrain – while the unadapted plants would perish and the poorly adapted plants would dwindle in number and be reduced to a small population in a less adverse ecological niche in the area.

Such a process of differentiation and selection was aided by spontaneous mutations in a population or by chance outcrossing between adjacent plants or both. The process could independently occur at many new sites of cultivation and recur when environmental conditions or cultivation practices changed. Therefore, rich genetic diversity of a secondary nature could be found in areas of undulating terrain where the environmental conditions significantly differed within a small area. The Assam and Madhya Pradesh states and Jeypore tract of India, the island of Sri Lanka, and Yunnan Province of China represent such areas of remarkable varietal diversity (Chang 1985).

Proliferation into Ecogeographic Races and Ecotypes

Continuous cultivation and intense selection in areas outside the conventional wetlands of shallow water depth (the paddies) have resulted in a range of extreme ecotypes: deepwater or floating rices that can cope with gradually rising waters up to 5 meters (m) deep; flood-tolerant rices that can survive days of total submergence under water; and upland or hill rices that are grown under dryland conditions like corn and sorghum. The varying soil-water-temperature regimes in the Bengal Bay states of India and in Bangladesh resulted in four seasonal ecotypes in that area: *boro* (winter), *aus* (summer), transplanted *aman* (fall, shallow water), and broadcast *aman* (fall, deep water). In many double-cropping areas, two main ecotypes follow the respective cropping season: dry (or off) and wet (or main) (Chang 1985).

In broader terms, the wide dispersal of *O. sativa* and subsequent isolation or selection in Asia has led to the formation of three ecogeographic races that differ in morphological and physiological characteristics and are partially incompatible in genetic affinity: Indica race in the tropics and subtropics; javanica race in the tropics; and sinica (or japonica) race in the temperate zone. Of the three races, indica is the oldest and the prototype of the other two races as it retains most of the primitive features: tallness, weak stems, lateness, dormant grains, and shattering panicles.

The sinica race became differentiated in China and has been rigorously selected for tolerance to cool temperatures, high productivity, and adaptiveness to modern cultivation technology: short plant stature, nitrogen responsiveness, earliness, stiff stems, and high grain yield. The javanica race is of more recent origin and appears intermediate between the other two races in genetic affinity, meaning it is more cross-fertile with either indica or sinica. Javanica cultivars are marked by gigas features in plant panicle and grain characters. They include a wetland group of cultivars (*bulu* and *gundil* of Indonesia) and a dryland group (hill rices of Southeast Asia).

The picture of race-forming processes is yet incomplete (Chang 1985). Many studies have relied heavily on grain size and shape as empirical criteria for race classification. Some studies employed crossing experiments and hybrid fertility ratings. Other workers recently used isozyme patterns to indicate origin and affinity. Controversies in past studies stemmed largely from limited samples, oversimplified empirical tests, and reliance on presently grown cultivars to retrace the distant past. The latter involved a lack of appreciation for the relatively short period (approximately 5 to 6 centuries) that it takes for a predominant grain type to be replaced by another (Watabe 1973), which was probably affected by the cultivator's preference. Most of the studies have also overlooked the usefulness of including amylose content and low temperature tolerance in revealing race identity (Chang 1976b, 1985). It should also be recognized that early human contacts greatly predated those given in historical records (Chang 1983), and maximum varietal diversity often showed up in places outside the area of primary genetic diversity (Chang 1976b, 1985).

Parallel to the expansion in production area and dispersal of the cultivars to new lands during the last two centuries was the growth of varietal diversity. In the first half of the twentieth century, before scientifically bred cultivars appeared in large numbers, the total number of unimproved varieties grown by Asian farmers probably exceeded 100,000, though many duplicates of similar or altered names were included in this tally (Chang 1984 and 1992).

The Spread of Asian Rice

Historical records are quite revealing of the spread of Asian rice from South Asia, Southeast Asia, and China to other regions or countries, though exact dates may be lacking. In the northward direction, the Sinica race was introduced from China into the Korean peninsula before 1030 B.C. (Chen 1989). Rice cultivation in Japan began in the late Jomon period (about 1000 B.C., [Akazawa 1983]), while earlier estimates placed the introduction of rice to Japan from China in the third century B.C. (Ando 1951; Morinaga 1968). Several routes could have been involved: (1) from the lower Yangtze basin to Kyushu island, (2)

from north China to Honshu Island, or (3) via Korea to northern Kyushu; *hsien* (*Indica*) may have arrived from China, and the Javanica race traveled from Southeast Asia (Isao 1976; Lu and Chang 1980). The areas that comprised the former Soviet Union obtained rice seeds from China, Korea, Japan, and Persia, and rice was grown around the Caspian Sea beginning in the early 1770s (Lu and Chang 1980).

From the Indian subcontinent and mainland Southeast Asia, the *Indica* race spread southward into Sri Lanka (before 543 B.C.), the Malay Archipelago (date unknown), the Indonesian islands (between 2000 and 1400 B.C.), and central and coastal China south of the Yangtze River. *Hsien* or *Indica*-type grains were found at both He-mu-du and Luo-jia-jiao sites in east China around 5000 B.C. (Lu and Chang 1980; Chang 1988). The *keng* or *sinica* rice was likely to have differentiated in the Yunnan-Kweichow region, and they became fixed in the cooler northern areas (Chang 1976b). On the other hand, several Chinese scholars maintain that *hsien* and *keng* rice were differentiated from wild rice inside China (Ting 1961; Yan 1989). The large-scale introduction and planting of the Champa rice (initially from Vietnam) greatly altered the varietal composition of *hsien* rice in south China and the central Yangtze basin after the eleventh century (Ho 1956; Chang 1987).

The *javanica* race had its origin on the Asian mainland before it differentiated into the dryland ecotype (related to the *aus* type of the Bengal Bay area and the hill rice of Southeast Asia) and the wetland ecotype (*bulu* and *gundil*) of Indonesia. From Indonesia, the wetland ecotype spread to the Philippines (mainly in the Ifugao region at about 1000 B.C.), Taiwan (at 2000 B.C. or later), and probably Ryukyu and Japan (Chang 1976b, 1988).

The Middle East acquired rice from South Asia probably as early as 1000 B.C. Persia loomed large as the principal stepping stone from tropical Asia toward points west of the Persian Empire. The Romans learned about rice during the expedition of Alexander the Great to India (c. 327–4 B.C.) but imported rice wine instead of growing the crop. The introduction of rice into Europe could have taken different routes: (1) from Persia to Egypt between the fourth and the first centuries B.C., (2) from Greece or Egypt to Spain and Sicily in the eighth century A.D., and (3) from Persia to Spain in the eighth century and later to Italy between the thirteenth and sixteenth centuries. The Turks brought rice from Southwest Asia into the Balkan Peninsula, and Italy could also have served as a stepping stone for rice growing in that region. Direct imports from various parts of Asia into Europe are also probable (Lu and Chang 1980).

In the spread of rice to Africa, Madagascar received Asian rice probably as early as 1000 B.C. when the early settlers arrived in the southwest region. Indonesian settlers who reached the island after the beginning of the Christian era brought in some Javanica

rices. Madagascar also served as the intermediary for the countries in East Africa, although direct imports from South Asia would have been another source. Countries in West Africa obtained Asian rice through European colonizers between the fifteenth and seventeenth centuries. Rice was also brought into Congo from Mozambique in the nineteenth century (Lu and Chang 1980).

The Caribbean islands obtained their rice from Europe in the late fifteenth and early sixteenth centuries. Central and South America received rice seeds from European countries, particularly Spain, during the sixteenth through the eighteenth centuries. In addition, there was much exchange of cultivars among countries of Central, South, and North America (Lu and Chang 1980).

Rice cultivation in the United States began around 1609 as a trial planting in Virginia. Other plantings soon followed along the south Atlantic coast. Rice production was well established in South Carolina by about 1690. It then spread to the areas comprising Mississippi and southwest Louisiana, to adjoining areas in Texas, and to central Arkansas, which are now the main rice-producing states in the South. California began rice growing in 1909–12 with the predominant cultivar the *sinica* type, which can tolerate cold water at the seedling stage. Rice was introduced into Hawaii by Chinese immigrants between 1853 and 1862, but it did not thrive as an agro-industry in competition with sugarcane and pineapple (Adair, Miller, and Beachell 1962; Lu and Chang 1980).

Experimental planting of rice in Australia took place in New South Wales in 1892, although other introductions into the warmer areas of Queensland and the Northern Territories could have come earlier. Commercial planting in New South Wales began in 1923 (Grist 1975). The island of New Guinea began growing rice in the nineteenth century (Bertin et al. 1971).

The dissemination of Asian rice from one place to another doubtless also took place for serendipitous reasons. Mexico, for example, received its first lot of rice seed around 1522 in a cargo mixed with wheat. South Carolina's early plantings of rice around 1685–94 allegedly used rice salvaged from a wrecked ship whose last voyage began in Madagascar (Grist 1975; Lu and Chang 1980).

In addition, the deliberate introduction of rice has produced other unexpected benefits. This occurred when the Champa rice of central Vietnam were initially brought to the coastal areas of South China. In 1011–12 the Emperor Chen-Tsung of the Sung Dynasty decreed the shipment of 30,000 bushels of seed from Fukien Province into the lower Yangtze basin because of the grain's early maturing and drought-escaping characteristics. But its subsequent widespread use in China paved the way for the double cropping of rice and the multiple cropping of rice and other crops (Ho 1956; Chang 1987).

As for African rice (*O. glaberrima*), its cultivation remains confined to West Africa under a variety of soil-water regimes: deep water basins, water holes in the savannas, hydromorphic soils in the forest zone, and dryland conditions in hilly areas (Porteres 1956; Harlan 1973). In areas favorable for irrigated rice production, African rice has been rapidly displaced by the Asian introductions, and in such fields the native cultigen has become a weed in commercial plantings.

It is interesting to note that the African cultigen has been found as far afield as Central America, most likely as a result of introduction during the time of the transatlantic slave trade (Bertin et al. 1971).

Cultivation Practices and Cultural Exchanges

Evolution of Cultivation Practices

Rice grains were initially gathered and consumed by prehistoric peoples in the humid tropics and subtropics from self-propagating wild stands. Cultivation began when men or, more likely, women, deliberately dropped rice grains on the soil in low-lying spots near their homesteads, kept out the weeds and animals, and manipulated the water supply. The association between rice and human community was clearly indicated in the exciting excavations at He-mu-du, Luo-jia-jiao, and Pen-tou-shan in China where rice was a principal food plant in the developing human settlements there more than 7,000 years ago.

Rice first entered the diet as a supplement to other food plants as well as to game, fish, and shellfish. As rice cultivation expanded and became more efficient, it replaced other cereals (millets, sorghums, Job's tears, and even wheat), root crops, and forage plants. The continuous expansion of rice cultivation owed much to its unique features as a self-supporting semiaquatic plant. These features include the ability of seed to germinate under both aerobic and anaerobic conditions and the series of air-conducting aerenchymatous tissues in the leafsheaths, stems, and roots that supply air to roots under continuous flooding. Also important are soil microbes in the root zone that fix nitrogen to feed rice growth, and the wide adaptability of rice to both wetland and dryland soil-water regimes. It is for these reasons that rice is the only subsistence crop whose soil is poorly drained and needs no nitrogen fertilizer applied. And these factors, in turn, account for the broad rice-growing belt from the Sino-Russian border along the Amur River (53°N latitude) to central Argentina (35°S).

Forces crucial to the expansion and improvement of rice cultivation were water control, farm implements, draft animals, planting methods, weed and pest control, manuring, seed selection, postharvest facilities, and above all, human innovation. A number of significant events selected from the voluminous historical records on rice are summarized below to illustrate the concurrent progression in its cultivation tech-

niques and the socio-politico-economic changes that accompanied this progression.

Rice was initially grown as a rain-fed crop in low-lying areas where rain water could be retained. Such areas were located in marshy, but flood-free, sites around river bends, as found in Honan and Shenxi Provinces of north China (Ho 1956), and at larger sites between small rivers, as represented by the He-mu-du site in east China (Chang 1968; You 1976). Early community efforts led to irrigation or drainage projects. The earliest of such activities in the historical record were flood-control efforts in the Yellow River area under Emperor Yu at about 2000 B.C. Irrigation works, including dams, canals, conduits, sluices, and ponds, were in operation during the Yin period (c. 1400 B.C.).

A system of irrigation and drainage projects of various sizes were set up during the Chou Dynasty. Large-scale irrigation works were built during the Warring States period (770–21 B.C.). By 400 B.C., "rice [*tao*] men" were appointed to supervise the planting and water management operations. The famous Tu-Cheng-Yen Dam was constructed near Chendu in Sichuan (Szechuan) Province about 250 B.C., which made western Sichuan the new rice granary of China.

Further developments during the Tang and Sung dynasties led to extensive construction of ponds as water reservoirs and of dams in a serial order to impound fresh water in rivers during high tides. Dykes were built around lake shores to make use of the rich alluvial soil (Chou 1986), and the importance of water quality was recognized (Amano 1979).

Among farm implements, tools made from stone (spade, hoe, axe, knife, grinder, pestle, and mortar) preceded those made from wood and large animal bones (hoe, spade); these were followed by bronze and iron tools. Bone spades along with wooden handles were found at the He-mu-du site. Bronze knives and sickles appeared during Shang and Western Chou. Between 770 and 211 B.C. iron tools appeared in many forms. The iron plow pulled by oxen was perfected during the Western Han period. Deep plowing was advocated from the third century B.C. onward. The spike-tooth harrow (*pa*) appeared around the Tang Dynasty (sixth century), and it markedly improved the puddling of wet soil and facilitated the transplanting process. This implement later spread to Southeast Asia to become an essential component in facilitating transplanted rice culture there (Chang 1976a). Other implements, such as the roller and a spiked board, were also developed to improve further the puddling and leveling operations.

Broadcasting rice grains into a low-lying site was the earliest method of planting and can still be seen in the Jeypore tract of India and many parts of Africa. In dry soils, the next development was to break through the soil with implements, mainly the plow, whereas in wetland culture, it was to build levees (short dikes or bunds) around a field in order to

impound the water. In the latter case, such an operation also facilitated land leveling and soil preparation by puddling the wet soil in repeated rounds.

The next giant step came in the transplanting (insertion) of young rice seedlings into a well-puddled and leveled wet field. Transplanting requires the raising of seedlings in nursery beds, then pulling them from those beds, bundling them, and transporting them to the field where the seedlings are thrust by hand into the softened wet soil. A well-performed transplanting operation also requires seed selection, the soaking of seeds prior to their initial sowing, careful management of the nursery beds, and proper control of water in the nursery and in the field. The transplanting practice began in the late Han period (A.D. 23–270) and subsequently spread to neighboring countries in Southeast Asia as a package comprised of the water buffalo, plow, and the spike-tooth harrow.

Transplanting is a labor-consuming operation. Depending on the circumstances, between 12 and close to 50 days of an individual's labor is required to transplant one hectare of rice land (Barker, Herdt, and Rose 1985). On the other hand, transplanting provides definite advantages in terms of a fuller use of the available water, especially during dry years, better weed control, more uniform maturation of the plants, higher grain yield under intensive management, and more efficient use of the land for rice and other crops in cropping sequence.

Despite these advantages, however, in South Asia the transplanting method remains second in popularity to direct seeding (broadcasting or drilling) due to operational difficulties having to do with farm implements, water control, and labor supply (Chang 1976a).

Variations of the one-step transplanting method were (1) to interplant an early maturing variety and a late one in alternating rows in two steps (once practiced in central China) and (2) to pull two-week-old seedlings as clumps and set them in a second nursery until they were about one meter tall. At this point, they were divided into smaller bunches and once more transplanted into the main field. This method, called double transplanting, is still practiced in Indochina in anticipation of quickly rising flood waters and a long rain season (Grist 1975).

Weeds in rice fields have undoubtedly been a serious production constraint since ancient times. The importance of removing weeds and wild rice plants was emphasized as early as the Han Dynasty. Widely practiced methods of controlling unwanted plants in the southern regions involved burning weeds prior to plowing and pulling them afterward, complemented by maintaining proper water depth in the field. Fallowing was mentioned as another means of weed control, and midseason drainage and tillage has been practiced since Eastern Chou as an effective means of weed control and of the suppression of late tiller formation by the rice plant.

Different tools, mainly of the hoe and harrow types, were developed for tillage and weed destruction. Otherwise, manual scratching of the soil surface and removal of weeds by hand were practiced by weederers who crawled forward among rows of growing rice plants. Short bamboo tubes tipped with iron claws were placed on the finger tips to help in the tedious operation. More efficient tools, one of which was a canoe-shaped wooden frame with a long handle and rows of spikes beneath it, appeared later (Amano 1979: 403). This was surpassed only by the rotary weeder of the twentieth century (Grist 1975: 157).

Insect pests were mentioned in Chinese documents before plant diseases were recognized. The *Odes* (c. sixth century B.C.) mentioned stemborers and the granary moth. During the Sung Dynasty, giant bamboo combs were used to remove leaf rollers that infest the upper portions of rice leaves. A mixture of lime and tung oil was used as an insect spray. Kernel smut, blast disease, and cold injury during flowering were recognized at the time of the Ming Dynasty. Seedling rot was mentioned in the *Agricultural Manual* of Chen Fu during South Sung (Chinese Academy of Agricultural Sciences 1986).

The relationship between manuring and increased rice yield was observed and recorded more than two thousand years ago. The use of compost and plant ash was advocated in writings of the first and third centuries. Boiling of animal bones in water as a means to extract phosphorus was practiced in Eastern Han. Growing a green manuring crop in winter was advised in the third century. The sixth century agricultural encyclopedia *Ch'i-Min-Yao-Shu* (Ku undated) distinguished between basal and top dressings of manure, preached the use of human and animal excreta on poor soils, and provided crop rotation schemes (Chang 1979b).

Irrigation practices received much attention in China because of the poor or erratic water supply in many rice areas. Therefore, the labor inputs on water management in Wushih County of Jiangsu Province in the 1920s surpassed those of weeding or transplanting by a factor of two (Amano 1979: 410), whereas in monsoonal Java, the inputs in water management were insignificant (Barker et al. 1985: 126).

Because of the cooler weather in north China, irrigation practices were attuned to suitable weather conditions as early as the Western Han: Water inlets and outlets were positioned directly opposite across the field so as to warm the flowing water by sunlight during the early stages of rice growth. Elsewhere, the inlets and outlets were repositioned at different intervals in order to cool the water during hot summer months (Amano 1979: 182). The encyclopedia *Ch'i-Min-Yao-Shu* devoted much space to irrigation practices: Watering should be attuned to the weather; the fields should be drained after tillage so as to firm the roots and drained again before harvesting.

In order to supplement the unreliable rainfall, many implements were developed to irrigate individual fields. The developments began with the use of urns to carry water from creeks or wells. The urn or bucket was later fastened to the end of a long pole and counterbalanced on the other end by a large chunk of stone. The pole was rested on a stand and could be swung around to facilitate the filling or pouring. A winch was later used to haul a bucket from a well (see Amano 1979 for illustrations).

The square-pallet chain pump came into use during the Eastern Han; it was either manually driven by foot pedaling or driven by a draft animal turning a large wheel and a geared transmission device (Amano 1979: 205, 240). The chain pump was extensively used in China until it was replaced by engine-driven water pumps after the 1930s. The device also spread to Vietnam. During hot and dry summers, the pumping operation required days and nights of continuous input. Other implements, such as the water wheel in various forms, were also used (Amano 1979; Chao 1979).

Although deepwater rice culture in China never approached the scale found in tropical Asia, Chinese farmers used floating rafts made of wooden frames and tied to the shore so as to grow rice in swamps. Such a practice appeared in Late Han, and the rafts were called *feng* (for frames) fields (Amano 1979: 175).

Many rice cultivars are capable of producing new tillers and panicles from the stubble after a harvest. Such regrowth from the cut stalks is called a ratoon crop. Ratooning was practiced in China as early as the Eastern Tsin period (A.D. 317–417), and it is now an important practice in the southern United States. Ratooning gives better returns in the temperate zone than in the tropics because the insects and diseases that persist from crop to crop pose more serious problems in the tropics.

Seed selection has served as a powerful force in cultivar formation and domestication. Continued selection by rice farmers in the field was even more powerful in fixing new forms; they used the desirable gene-combinations showing up in the plantings to suit their farmer's different needs and fancies. The earliest mention of human-directed selection in Chinese records during the first century B.C. was focused on selecting panicles with more grains and fully developed kernels. Soon, varietal differences in awn color and length, maturity, grain size and shape, stickiness of cooked rice, aroma of milled rice, and adaptiveness to dryland farming were recognized. The trend in selection was largely toward an earlier maturity, which reduced cold damage and made multiple cropping more practical in many areas. The encyclopedia *Ch'i-Min-Yao-Shu* advised farmers to grow seeds in a separate plot, rotate the seed plot site in order to eliminate weedy rice, and select pure and uniformly colored panicles. The seeds were to be stored above ground in aerated baskets, not under the ground. Seed selection by winnowing and floatation in water was advised.

Dryland or hill rice was mentioned in writings of the third century B.C. (Ting 1961). During Eastern Tsin, thirteen varieties were mentioned; their names indicated differences in pigmentation of awn, stem and hull, maturity, grain length, and stickiness of cooked rice (Amano 1979). Varieties with outstanding grain quality frequently appeared in later records. Indeed, a total of 3,000 varieties was tallied, and the names were a further indication of the differences in plant stature and morphology, panicle morphology, response to manuring, resistance to pests, tolerance to stress factors (drought, salinity, alkalinity, cool temperatures, and deep water), and ratooning ability (You 1982). The broad genetic spectrum present in the rice varieties of China was clearly indicated.

Harvesting and processing rice is another laborious process. The cutting instruments evolved from knives to sickles to scythe. Community efforts were common in irrigated areas, and such neighborly cooperation can still be seen in China, Indonesia, the Philippines, Thailand, and other countries. Threshing of grains from the panicles had been done in a variety of ways: beating the bundle of cut stalks against a wooden bench or block; trampling by human feet or animal hoofs; beating with a flail; and, more recently, driving the panicles through a spiked drum that is a prototype of the modern grain combine (see Amano 1979: 248–54 for the ancient tools).

Other important postharvest operations are the drying of the grain (mainly by sun drying), winnowing (by natural breeze or a hand-cranked fan inside a drum winnower), dehulling (dehulling), and milling (by pestle and mortar, stone mills, or modern dehulling and milling machines). Grains and milled rice are stored in sacks or in bulk inside bins. In Indonesia and other hilly areas, the long-panicked Javanica rices are tied into bundles prior to storage.

To sum up the evolutionary pathway in wetland rice cultivation on a worldwide scale, cultivation began with broadcasting in rain-fed and unbunded fields under shifting cultivation. As the growers settled down, the cultivation sites became permanent fields. Then, bunds were built to impound the rain water, and the transplanting method followed. As population pressure on the land continued to increase, irrigation and transplanting became more imperative (Chang 1989a).

The entire range of practices can still be seen in the Jeypore Tract and the neighboring areas (author's personal observations). The same process was retraced in Bang Chan (near Bangkok) within a span of one hundred years. In this case, the interrelationships among land availability, types of rice culture, population density, labor inputs, and grain outputs were documented in a fascinating book entitled *Rice and Man* by L. M. Hanks (1972).

In the twentieth century, further advances in agricultural engineering and technology have to do with several variations in seeding practices that have been

adopted to replace transplanting. Rice growers in the southern United States drill seed into a dry soil. The field is briefly flushed with water and then drained. The seeds are allowed to germinate, and water is introduced when the seedlings are established. In northern California, pregerminated seeds are dropped from airplanes into cool water several inches deep. The locally selected varieties are able to emerge from the harsh environment (Adair et al. 1973).

Recently, many Japanese farmers have turned to drill-plant pregerminated seed on wet mud. An oxidant is applied to the seed before sowing so as to obtain a uniform stand of plants. For the transplanted crop, transplanting machines have been developed not only to facilitate this process but also to make commercial raising of rice seedlings inside seed boxes a profitable venture. As labor costs continue to rise worldwide, direct seeding coupled with chemical weed control will be the main procedures in the future.

For deepwater rice culture, rice seeds are broadcast on dry soil. The seeds germinate after the monsoon rains arrive. The crop is harvested after the rains stop and the flooding water has receded.

For dryland rice, seeds are either broadcasted, drilled, or dropped (dibbled) into shallow holes dug in the ground. Dibbling is also common in West Africa. Dryland (hill or upland) rice continues to diminish in area because of low and unstable yield. It has receded largely into hilly areas in Asia where tribal minorities and people practicing shifting cultivation grow small patches for subsistence.

Rice Cultivation and Cultural Exchanges

The expansion of rice cultivation in China involved interactions and exchanges in cultural developments, human migration, and progress in agricultural technology. Agricultural technology in north China developed ahead of other regions of China. Areas south of the Yangtze River, especially south China, were generally regarded by Chinese scholars of the north as primitive in agricultural practices. During travel to the far south in the twelfth century, one of these scholars described the local rain-fed rice culture. He regarded it as crude in land preparation: Seed was sown by dibbling, fertilizer was not used, and tillage as a weeding practice was unknown (Ho 1969).

However, the picture has been rather different in the middle and lower Yangtze basins since the Tsin Dynasty (beginning in A.D. 317) when a mass migration of people from the north to southern areas took place. The rapid expansion of rice cultivation in east China was aided by the large-scale production of iron tools used in clearing forests and the widespread adoption of transplanting.

Private land ownership, which began in the Sung (beginning in A.D. 960), followed by reduction of land rent in the eleventh century and reinforced by double cropping and growth in irrigation works, stimulated

rice production and technology development. As a result, rice production south of the Yangtze greatly surpassed rice production in the north, and human population growth followed the same trend (Ho 1969; Chang 1987). Thus, the flow of rice germ plasm was from south to north, but much of the cultural and technological developments diffused in the opposite direction.

Culinary Usage and Nutritional Aspects

Rice Foods

Before the rice grain is consumed, the silica-rich husk (hull, chaff) must be removed. The remaining kernel is the caryopsis or brown rice. Rice consumers, however, generally prefer to eat milled rice, which is the product after the bran (embryo and various layers of seed coat) is removed by milling. Milled rice is, invariably, the white, starchy endosperm, despite pigments present in the hull (straw, gold, brown, red, purple or black) and in the seed coat (red or purple).

Parboiled rice is another form of milled rice in which the starch is gelatinized after the grain is pre-cooked by soaking and heating (boiling, steaming, or dry heating), followed by drying and milling. Milled rice may also be ground into a powder (flour), which enters the food industry in the form of cakes, noodles, baked products, pudding, snack foods, infant formula, fermented items, and other industrial products.

Fermentation of milled glutinous rice or overmilled nonglutinous rice produces rice wine (*sake*). Vinegar is made from milled and broken rice and beer from broken rice and malt. Although brown rice, as well as lightly milled rice retaining a portion of the germ (embryo), are recommended by health-food enthusiasts, their consumption remains light. Brown rice is difficult to digest due to its high fiber content, and it tends to become rancid during extended storage. Cooking of all categories of rice is done by applying heat (boiling or steaming) to soaked rice until the kernels are fully gelatinized and excess water is expelled from the cooked product. Cooked rice can be lightly fried in oil to make fried rice. People of the Middle East prefer to fry the rice lightly before boiling. Americans often add salt and butter or margarine to soaked rice prior to boiling. The peoples of Southeast Asia eat boiled rice three times a day, including breakfast, whereas peoples of China, Japan, and Korea prepare their breakfast by boiling rice with excess water, resulting in porridge (thick gruel) or *congee* (thin soup).

Different kinds of cooked rice are distinguished by cohesiveness or dryness, tenderness or hardness, whiteness or other colors, flavor or taste, appearance, and aroma (or its absence). Of these features, cohesiveness or dryness is the most important varietal characteristic: High amylose (25 to 30 percent) of the starchy endosperm results in dry and fluffy kernels; intermediate amylose content (15 to 25 percent) produces tender and slightly cohesive rice; low amylose

content (10 to 15 percent) leads to soft cohesive (aggregated) rice; and glutinous or waxy endosperm (0.8 to 1.3 percent amylose) produces highly sticky rice. Amylopectin is the other – and the major – fraction of rice starch in the endosperm.

These four classes of amylose content and cooked products largely correspond with the designation of *Indica*, *Javanica*, *Sinica* (*Japonica*), and glutinous. Other than amylose content, the cooked rice is affected by the rice-water ratio, cooking time, and age of rice. Hardness, flavor, color, aroma, and texture of the cooked rice upon cooling are also varietal characteristics (Chang 1988; Chang and Li 1991).

Consumer preference for cooked rice and other rice products varies greatly from region to region and is largely a matter of personal preference based on upbringing. For instance, most residents of Shanghai prefer the cohesive *keng* (*Sinica*) rice, whereas people in Nanjing about 270 kilometers away in the same province prefer the drier *hsien* (*Indica*) type. Tribal people of Burma, Laos, Thailand, and Vietnam eat glutinous rice three times a day – a habit unthinkable to the people on the plains. Indians and Pakistanis pay a higher price for the basmati rices, which elongate markedly upon cooking and have a strong aroma. People of South Asia generally prefer slender-shaped rice, but many Sri Lankans fancy the short, roundish *samba* rices, which also have dark red seed coats. Red rice is also prized by tribal people of Southeast Asia (Eggum et al. 1981; Juliano 1985c) and by numerous Asians during festivities, but its alleged nutritional advantage over ordinary rice remains a myth. It appears that the eye appeal of red or purple rice stems from the symbolic meaning given the color red throughout Asia, which is “good luck.”

The pestle and mortar were doubtless the earliest implements used to mill rice grains. The milling machines of more recent origin use rollers that progressed from stone to wood to steel and then to rubber-wrapped steel cylinders. Tubes made of sections of bamboo were most likely an early cooking utensil, especially for travelers. A steamer made of clay was unearthed at the He-mu-du site dating from 5000 B.C., but the ceramic and bronze pots were the main cooking utensils until ironware came into use. Electric rice cookers replaced iron or aluminum pots in Japan and other Asian countries after the 1950s, and today microwave ovens are used to some extent.

Nutritional Considerations

Rice is unquestionably a superior source of energy among the cereals. The protein quality of rice (66 percent) ranks only below that of oats (68 percent) and surpasses that of whole wheat (53 percent) and of corn (49 percent). Milling of brown rice into white rice results in a nearly 50 percent loss of the vitamin B complex and iron, and washing milled rice prior to cooking further reduces the water-soluble vitamin content. However, the amino acids, especially lysine, are less

affected by the milling process (Kik 1957; Mickus and Luh 1980; Juliano 1985a; Juliano and Bechtel 1985).

Rice, which is low in sodium and fat and is free of cholesterol, serves as an aid in treating hypertension. It is also free from allergens and now widely used in baby foods (James and McCaskill 1983). Rice starch can also serve as a substitute for glucose in oral rehydration solution for infants suffering from diarrhea (Juliano 1985b).

The development of beriberi by people whose diets have centered too closely on rice led to efforts in the 1950s to enrich polished rice with physiologically active and rinse-free vitamin derivatives. However, widespread application was hampered by increased cost and yellowing of the kernels upon cooking (Mickus and Luh 1980). Certain states in the United States required milled rice to be sold in an enriched form, but the campaign did not gain acceptance in the developing countries. After the 1950s, nutritional intakes of the masses in Asia generally improved and, with dietary diversification, beriberi receded as a serious threat.

Another factor in keeping beriberi at bay has been the technique of parboiling rough rice. This permits the water-soluble vitamins and mineral salts to spread through the endosperm and the proteinaceous material to sink into the compact mass of gelatinized starch. The result is a smaller loss of vitamins, minerals, and amino acids during the milling of parboiled grains (Mickus and Luh 1980), although the mechanism has not been fully understood. Parboiled rice is popular among the low-income people of Bangladesh, India, Nepal, Pakistan, Sri Lanka, and parts of West Africa and amounts to nearly one-fifth of the world's rice consumed (Bhattacharya 1985).

During the 1970s, several institutions attempted to improve brown rice protein content by breeding. Unfortunately, such efforts were not rewarding because the protein content of a variety is highly variable and markedly affected by environment and fertilizers, and protein levels are inversely related to levels of grain yield (Juliano and Bechtel 1985).

Production and Improvement in the Twentieth Century

Production Trends

Prior to the end of World War II, statistical information on global rice production was rather limited in scope. The United States Department of Agriculture (USDA) compiled agricultural statistics in the 1930s, and the Food and Agriculture Organization of the United Nations (FAO) expanded these efforts in the early 1950s (FAO 1965). In recent years, the *World Rice Statistics* published periodically by the International Rice Research Institute (IRRI) provides comprehensive information on production aspects, imports and exports, prices, and other useful information concerning rice (IRRI 1991).

During the first half of the twentieth century, production growth stemmed largely from an increase in wetland rice area and, to a lesser extent, from expansion of irrigated area and from yields increased by the use of nitrogen fertilizer. Then, varietal improvement came in as the vehicle for delivering higher grain yields, especially in the late 1960s when the "Green Revolution" in rice began to gather momentum (Chang 1979a).

Rice production in Asian countries steadily increased from 240 million metric tons during 1964–6 to 474 million tons in 1989–90 (IRRI 1991). Among the factors were expansion in rice area and/or irrigated area; adoption of high-yielding, semidwarf varieties (HYVs); use of nitrogen fertilizers and other chemicals (insecticides, herbicides, and fungicides); improved cultural methods; and intensified land use through multiple cropping (Herd and Capule 1983; Chang and Luh 1991).

Asian countries produced about 95 percent of the world's rice during the years 1911–40. After 1945, however, Asia's share dropped to about 92 percent by the 1980s, with production growth most notable in North and South America (IRRI 1991; information on changes in grain yield, production, annual growth rates, and prices in different Asian countries is provided in Chang 1993b; Chang and Luh 1991; David 1991; and Chang 1979a).

But despite the phenomenal rise in crop production and (in view of rapidly growing populations) the consequent postponement of massive food shortages in Asia since the middle 1960s, two important problems remain. One of these is food production per capita, which advanced only slightly ahead of population growth (WRI 1986). The other is grain yield, which remained low in adverse rain-fed environments – wetland, dryland, deepwater, and tidal swamps (IRRI 1989). In fact, an apparent plateau has prevailed for two decades in irrigated rice (Chang 1983). Moreover, the cost of fertilizers, other chemicals, labor, and good land continued to rise after the 1970s, whereas the domestic wholesale prices in real terms slumped in most tropical Asian nations and have remained below the 1966–8 level.

This combination of factors brought great concern when adverse weather struck many rice areas in Asia in 1987 and rice stocks became very low. Fortunately, weather conditions improved the following year and rice production rebounded (Chang and Luh 1991; IRRI 1991).

However, the threat to production remains. In East Asia, five years of favorable weather ended in 1994 with a greater-than-usual number of typhoons that brought massive rice shortages to Japan and South Korea. And in view of the "El Niño" phenomenon, a higher incidence of aberrant weather can be expected, which will mean droughts for some and floods for others (Nicholls 1993).

Germ Plasm Loss and the Perils of Varietal Uniformity

Rice is a self-fertilizing plant. Around 1920, however, Japanese and U.S. rice breeders took the lead in using scientific approaches (hybridization selection and testing) to improve rice varieties. Elsewhere, pureline selection among farmers' varieties was the main method of breeding.

After World War II, many Asian countries started to use hybridization as the main breeding approach. Through the sponsorship of the FAO, several countries in South and Southeast Asia joined in the Indica-Japonica Hybridization Project during the 1950s, exchanging rice germ plasm and using diverse parents in hybridization.

These efforts, however, provided very limited improvement in grain yield (Parthasarathy 1972), and the first real breakthrough came during the mid-1950s when Taiwan (first) and mainland China (second) independently succeeded in using their semidwarf rices in developing short-statured, nitrogen-responsive and high-yielding semidwarf varieties (HYVs). These HYVs spread quickly among Chinese rice farmers (Chang 1961; Huang, Chang, and Chang 1972; Shen 1980).

Taiwan's semidwarf "Taichung Native 1" (TN1) was introduced into India through the International Rice Research Institute (IRRI) located in the Philippines. "TN1" and IRRI-bred "IR8" triggered the "Green Revolution" in tropical rices (Chandler 1968; Huang et al. 1972). Subsequent developments in the dramatic spread of the HYVs and an associated rise in area grain yield and production have been documented (Chang 1979a; Dalrymple 1986), and refinements in breeding approaches and international collaboration have been described (Brady 1975; Khush 1984; Chang and Li 1991).

In the early 1970s, China scored another breakthrough in rice yield when a series of hybrid rices (F_1 hybrids) were developed by the use of a cytoplasmic pollen-sterile source found in a self-sterile wild plant ("Wild Abortive") on Hainan Island (Lin and Yuan 1980). The hybrids brought another yield increment (15 to 30 percent) over the widely grown semidwarfs.

Along with the rapid and large-scale adoption of the HYVs and with deforestation and development projects, innumerable farmers' traditional varieties of all three ecogenetic races and their wild relatives have disappeared from their original habitats – an irreversible process of "genetic erosion." The lowland group of the javanic race (*bulu*, *gundill*) suffered the heaviest losses on Java and Bali in Indonesia. Sizable plantings of the long-bearded *bulus* can now be found only in the Ifugao rice terraces of the Philippines.

In parallel developments, by the early 1990s the widespread planting of the semidwarf HYVs and hybrid rices in densely planted areas of Asia amounted to about 72 million hectares. These HYVs

share a common semidwarf gene (*sd₁*) and largely the same cytoplasm (either from "Cina" in older HYVs or "Wild Abortive" in the hybrids). This poses a serious threat of production losses due to a much narrowed genetic base if wide-ranging pest epidemics should break out, as was the case with hybrid maize in the United States during 1970-1 (Chang 1984).

Since the early 1970s, poorly educated rice farmers in South and Southeast Asia have planted the same HYV in successive crop seasons and have staggered plantings across two crops. Such a biologically unsound practice has led to the emergence of new and more virulent biotypes of insect pests and disease pathogens that have overcome the resistance genes in the newly bred and widely grown HYVs. The result has been heavy crop losses in several tropical countries in a cyclic pattern (Chang and Li 1991; Chang 1994).

Fortunately for the rice-growing world, the IRRI has, since its inception, assembled a huge germ plasm collection of more than 80,000 varieties and 1,500 wild rices by exchange and field collection. Seeds drawn from the collection not only have sustained the continuation of the "Green Revolution" in rice all over the world but also assure a rich reservoir of genetic material that can reinstate the broad genetic base in Asian rices that in earlier times kept pest damage to manageable levels (Chang 1984, 1989b, 1994).

Outlook for the Future

Since the dawn of civilization, rice has served humans as a life-giving cereal in the humid regions of Asia and, to a lesser extent, in West Africa. Introduction of rice into Europe and the Americas has led to its increased use in human diets. In more recent times, expansion in the rice areas of Asia and Africa has resulted in rice replacing other dryland cereals (including wheat) and root crops as the favorite among the food crops, wherever the masses can afford it. Moreover, a recent overview of food preferences in Africa, Latin America, and north China (Chang 1987, personal observation in China) suggests that it is unlikely that rice eaters will revert to such former staples as coarse grains and root crops. On the other hand, per capita rice consumption has markedly dropped in the affluent societies of Japan and Taiwan.

In the eastern half of Asia, where 90 to 95 percent of the rice produced is locally consumed, the grain is the largest source of total food energy. In the year 2000, about 40 percent of the people on earth, mostly those in the populous, less-developed countries, depended on rice as the major energy source. The question, of course, is whether the rice-producing countries with ongoing technological developments can keep production levels ahead of population growth.

From the preceding section on cultivation practices, it seems obvious that rice will continue to be a

labor-intensive crop on numerous small farms. Most of the rice farmers in rain-fed areas (nearly 50 percent of the total planted area) will remain subsistence farmers because of serious ecological and economic constraints and an inability to benefit from the scientific innovations that can upgrade land productivity (Chang 1993b). Production increases will continue to depend on the irrigated areas and the most favorable rain-fed wetlands, which now occupy a little over 50 percent of the harvested rice area but produce more than 70 percent of the crop. The irrigated land area may be expanded somewhat but at a slower rate and higher cost than earlier. Speaking to this point is a recent study that indicates that Southeast Asia and South Asia as well, are rapidly depleting their natural resources (Brookfield 1993).

With rising costs in labor, chemicals, fuel, and water, the farmers in irrigated areas will be squeezed between production costs and market price. The latter, dictated by government pricing policy in most countries, remains lower than the real rice price (David 1991). Meanwhile, urbanization and industrialization will continue to deprive the shrinking farming communities of skilled workers, especially young men. Such changes in rice-farming communities will have serious and widespread socioeconomic implications.

Unless rice farmers receive an equitable return for their efforts, newly developed technology will remain experimental in agricultural stations and colleges. The decision makers in government agencies and the rice-consuming public need to ensure that a decent living will result from the tilling of rice lands. Incentives must also be provided to keep skilled and experienced workers on the farms. Moreover, support for the agricultural research community must be sustained because the challenges of providing still more in productivity-related cultivation innovations for rice are unprecedented in scope.

Although the rice industry faces formidable challenges, there are areas that promise substantial gains in farm productivity with the existing technology of irrigated rice culture. A majority of rice farmers can upgrade their yields if they correctly and efficiently perform the essential cultivation practices of fertilization, weed and pest control, and water management.

On the research front, rewards can be gained by breaking the yield ceiling, making pest resistance more durable, and improving the tolerance to environmental stresses. Biotechnology will serve as a powerful force in broadening the use of exotic germ plasm in *Oryza* and related genera (Chang and Vaughan 1991). We also need the inspired and concerted teamwork of those various sectors of society that, during the 1960s and 1970s, made the "Green Revolution" an unprecedented event in the history of agriculture.

Lastly, control of human population, especially in the less-developed nations, is also crucial to the maintenance of an adequate food supply for all sectors of

human society. Scientific breakthroughs alone will not be able to relieve the overwhelming burden placed on the limited resources of the earth by uncontrolled population growth.

Te-Tzu Chang

Bibliography

- Adair, C. R., J. G. Atkins, C. N. Bollich, et al. 1973. *Rice in the United States: Varieties and production*. U.S. Department of Agriculture Handbook No. 289. Washington, D.C.
- Adair, C. R., M. D. Miller, and H. M. Beachell. 1962. Rice improvement and culture in the United States. *Advances in Agronomy* 14: 61-108.
- Akazawa, T. 1983. An outline of Japanese prehistory. In *Recent progress of natural sciences in Japan*, Vol. 8, *Anthropology*, 1-11. Tokyo.
- Alexander, J., and D. G. Coursey. 1969. The origins of yam cultivation. In *The domestication and exploitation of plants and animals*, ed. P. J. Ucko and G. W. Dimbleby, 323-9. London.
- Amano, M. 1979. *Chinese agricultural history research*. Revised edition (in Japanese). Tokyo.
- Andersson, J. G. 1934. *Children of the yellow earth: Studies in prehistoric China*. London.
- Ando, H. 1951. *Miscellaneous records on the ancient history of rice crop in Japan* (in Japanese). Tokyo.
- Andrus, J. R., and A. F. Mohammed. 1958. *The economy of Pakistan*. Oxford.
- Barker, R., R. W. Herdt, and B. Rose. 1985. *The rice economy of Asia*. Washington, D.C.
- Bertin, J., J. Hermardinquer, M. Keul, et al. 1971. *Atlas of food crops*. Paris.
- Bhattacharya, K. R. 1985. Parboiling of rice. In *Rice: Chemistry and technology*, ed. B. O. Juliano, 289-348. St. Paul, Minn.
- Brady, N. C. 1975. Rice responds to science. In *Crop productivity - research imperatives*, ed. A. W. A. Brown et al., 61-96. East Lansing, Mich.
- Brookfield, H. 1993. Conclusions and recommendations. In *South-East Asia's environmental future: The search for sustainability*, ed. H. Brookfield and Y. Byron, 363-73. Kuala Lumpur and Tokyo.
- Candolle, A. de. 1884. *Origin of cultivated plants* (1886 English translation). New York.
- Chandler, R. F., Jr. 1968. Dwarf rice - a giant in tropical Asia. In *Science for better living, U.S.D.A. 1968 Yearbook of Agriculture*, 252-5. Washington, D.C.
- Chang, K. C. 1968. *The archaeology of ancient China*. Revised edition. New Haven, Conn.
- Chang, T. T. 1961. Recent advances in rice breeding. In *Crop and seed improvement in Taiwan, Republic of China*, 33-58. Taipei.
- 1976a. The rice cultures. In *The early history of agriculture*, ed. J. Hutchinson et al., 143-55. London.
- 1976b. The origin, evolution, cultivation, dissemination, and divergence of Asian and African rices. *Euphytica* 25: 425-45.
- 1979a. Genetics and evolution of the Green Revolution. In *Replies from biological research*, ed. R. de Vicente, 187-209. Madrid.
- 1979b. History of early rice cultivation (in Chinese). In *Chinese agricultural history - collection of essays*, ed. T. H. Shen et al. Taipei.
1983. The origins and early cultures of the cereal grains and food legumes. In *The origins of Chinese civilization*, ed. D. N. Keightley, 65-94. Berkeley, Calif.
1984. Conservation of rice genetic resources: Luxury or necessity? *Science* 224: 251-6.
1985. Crop history and genetic conservation in rice - a case study. *Iowa State Journal of Research* 59: 405-55.
1987. The impact of rice in human civilization and population expansion. *Interdisciplinary Science Reviews* 12: 63-9.
1988. The ethnobotany of rice in island Southeast Asia. *Asian Perspectives* 26: 69-76.
- 1989a. Domestication and spread of the cultivated rices. In *Foraging and farming - the evolution of plant exploitation*, ed. D. R. Harris and G. C. Hillman, 408-17. London.
- 1989b. The management of rice genetic resources. *Genome* 31: 825-31.
1992. Availability of plant germplasm for use in crop improvement. In *Plant breeding in the 1990s*, ed. H. T. Stalker and J. P. Murphy, 17-35. Cambridge.
- 1993a. The role of germ plasm and genetics in meeting global rice needs. In *Advances in botany*, ed. Y. I. Hsing and C. H. Chou, 25-33. Taipei.
- 1993b. Sustaining and expanding the "Green Revolution" in rice. In *South-East Asia's environmental future: The search for sustainability*, ed. H. Brookfield and Y. Byron, 201-10. Kuala Lumpur and Tokyo.
1994. The biodiversity crisis in Asian crop production and remedial measures. In *Biodiversity and terrestrial ecosystems*, ed. C. I. Peng and C. H. Chou, 25-44. Taipei.
- Chang, T. T., and C. C. Li. 1991. Genetics and breeding. In *Rice*, Vol. 1, *Production*, ed. B. S. Luh, 23-101. Second edition. New York.
- Chang, T. T., and B. S. Luh. 1991. Overview and prospects of rice production. In *Rice*, Vol. 1, *Production*, ed. B. S. Luh, 1-11. Second edition. New York.
- Chang, T. T., and D. A. Vaughan. 1991. Conservation and potentials of rice genetic resources. In *Biotechnology in agriculture and forestry*, Vol. 14, *Rice*, ed. Y. P. S. Bajaj, 531-52. Berlin.
- Chao, Y. S. 1979. The Chinese water wheel (in Chinese). In *Chinese agricultural history*, ed. T. H. Shen and Y. S. Chao, 69-163. Taipei.
- Chatterjee, D. 1951. Note on the origin and distribution of wild and cultivated rices. *Indian Journal of Genetics and Plant Breeding* 11: 18-22.
- Chekiang Provincial Cultural Management Commission and Chekiang Provincial Museum. 1976. Ho-mu-tu discovery of important primitive society, an important remain (in Chinese). *Wen Wu* 8: 8-13.
- Chen, T. K., ed. 1958. *Rice* (part 1) (in Chinese). Peking.
- Chen, T. K. 1960. Lowland rice cultivation in Chinese literature (in Chinese). *Agricultural History Research Series* 2: 64-93.
- Chen, W. H. 1989. Several problems concerning the origin of rice growing in China (in Chinese). *Agricultural Archaeology* 2: 84-98.
- Chevalier, A. 1932. Nouvelle contribution à l'étude systématique des *Oryza*. *Review Botanique Appliquée et Agricola Tropical* 12: 1014-32.
- Chinese Academy of Agricultural Sciences. 1986. *Chinese rice science* (in Chinese). Beijing.

- Chou, K. Y. 1986. Farm irrigation of ancient China (in Chinese). *Agricultural Archaeology* 1: 175-83; 2: 168-79.
- Dalrymple, D. G. 1986. *Development and spread of high-yielding rice varieties in developing countries*. Washington, D.C.
- Dao, T. T. 1985. Types of rice cultivation and its related civilization in Vietnam. *East Asian Cultural Studies* 24: 41-56.
- David, C. C. 1991. The world rice economy: Challenges ahead. In *Rice biotechnology*, ed. G. S. Khush and G. Toennissen, 1-18. Cambridge.
- Eggum, B. O., E. P. Alabata, and B. O. Juliano. 1981. Protein utilization of pigmented and non-pigmented brown and milled rices by rats. *Quality of Plants, Plant Foods and Human Nutrition* 31: 175-9.
- FAO (Food and Agriculture Organization of the United Nations). 1965. *The world rice economy in figures, 1909-1963*. Rome.
- Grist, D. H. 1975. *Rice*. Fifth edition. London.
- Hanks, L. M. 1972. *Rice and man*. Chicago.
- Harlan, J. R. 1973. Genetic resources of some major field crops in Africa. In *Genetic resources in plants - their exploration and conservation*, ed. O. H. Frankel and E. Bennett, 19-32. Philadelphia, Pa.
- Hegsted, D. M. 1969. Nutritional value of cereal proteins in relation to human needs. In *Protein-enriched cereal foods for world needs*, ed. M. Milner, 38-48. St. Paul, Minn.
- Herd, R. W., and C. Capule. 1983. *Adoption, spread and production impact of modern rice varieties in Asia*. Los Baños, Philippines.
- Higham, C. F. W. 1989. Rice cultivation and the growth of Southeast Asian civilization. *Endeavour* 13: 82-8.
- Ho, P. T. 1956. Early ripening rice in Chinese history. *Economic History Review* 9: 200-18.
1969. The loess and the origin of Chinese agriculture. *American Historical Review* 75: 1-36.
- Hsia, N. 1977. Carbon-14 determined dates and prehistoric Chinese archaeological history (in Chinese). *K'ao-ku* 4: 217-32.
- Huang, C. H., W. L. Chang, and T. T. Chang. 1972. Ponlai varieties and Taichung Native 1. In *Rice breeding*, 31-46. Los Baños, Philippines.
- Hutchinson, J. 1976. India: Local and introduction crops. In *The early history of agriculture*, Philosophical Transactions of the Royal Society of London B275: 129-41.
- IRRI (International Rice Research Institute). 1989. *IRRI toward 2000 and beyond*. Los Baños, Philippines.
1991. *World rice statistics, 1990*. Los Baños, Philippines.
- Isao, H. 1976. *History of Japan as revealed by rice cultivation* (in Japanese). Tokyo.
- James, C., and D. McCaskill. 1983. Rice in the American diet. *Cereal Foods World* 28: 667-9.
- Juliano, B. O. 1985a. Production and utilization of rice. In *Rice: Chemistry and technology*, ed. B. O. Juliano, 1-16. St. Paul, Minn.
- 1985b. Polysaccharides, proteins and lipids of rice. In *Rice: Chemistry and technology*, ed. B. O. Juliano, 59-174. St. Paul, Minn.
- 1985c. Biochemical properties of rice. In *Rice: Chemistry and technology*, ed. B. O. Juliano, 175-205. St. Paul, Minn.
- Juliano, B. O., and D. B. Bechtel. 1985. The rice grain and its gross composition. In *Rice: Chemistry and technology*, ed. B. O. Juliano, 17-57. St. Paul, Minn.
- Khush, G. S. 1984. IRRI breeding program and its worldwide impact on increasing rice production. In *Genetic manipulation in plant improvement*, ed. J. P. Gustafson, 61-94. New York.
- Kik, M. C. 1957. The nutritive value of rice and its by-products. *Arkansas Agricultural Experiment Station Bulletin* 589.
- Ku, S. H. Undated. *Ch'i min yao shu*. Taipei.
- Lin, S. C., and L. P. Yuan. 1980. Hybrid rice breeding in China. In *Innovative approaches to rice breeding*, 35-52. Los Baños, Philippines.
- Lu, J. J., and T. T. Chang. 1980. Rice in its temporal and spatial perspectives. In *Rice: Production and utilization*, ed. B. S. Luh, 1-74. Westport, Conn.
- Melville, R. 1966. Mesozoic continents and the migration of the angiosperms. *Nature* 220: 116-20.
- Mickus, R. R., and B. S. Luh. 1980. Rice enrichment with vitamins and amino acids. In *Rice: Production and utilization*, ed. B. S. Luh, 486-500. Westport, Conn.
- Morinaga, T. 1951. *Rice of Japan* (in Japanese). Tokyo.
1968. Origin and geographical distribution of Japanese rice. *Japan Agricultural Research Quarterly* 3: 1-5.
- Nicholls, N. 1993. ENSO, drought, and flooding rain in South-East Asia. In *South-East Asia's environmental future: The search for sustainability*, ed. H. Brookfield and Y. Byron, 154-75. Kuala Lumpur and Tokyo.
- Oka, H. I. 1988. *Origin of cultivated rice*. Amsterdam and Tokyo.
- Pankar, S. N., and M. K. M. Gowda. 1976. On the origins of rice in India. *Science and Culture* 42: 547-50.
- Parthasarathy, N. 1972. Rice breeding in tropical Asia up to 1960. In *Rice breeding*, 5-29. Los Baños, Philippines.
- Pei, A. P. 1989. Rice remains in Peng-tou-shan culture in Hunan and rice growing in prehistorical China (in Chinese). *Agricultural Archaeology* 2: 102-8.
- Porteres, R. 1956. Taxonomie agrobotanique des riz cultives *O. sativa* Linn. et *O. glaberrima* Steud. *Journal Agricultural Tropical Botanique et Appliquée* 3: 341-84, 541-80, 627-70, 821-56.
- Randhawa, M. S. 1980. *A history of agriculture in India*, Vol. 1. New Delhi.
- Revel, N., ed. 1988. *Le riz en Asie du Sud-Est: Atlas du vocabulaire de la plante*. Paris.
- Roschevitz, R. J. 1931. A contribution to the knowledge of rice (in Russian with English summary). *Bulletin of Applied Botany, Genetics and Plant Breeding* (Leningrad) 27: 1-133.
- Sauer, C. O. 1952. *Agricultural origin and dispersals*. New York.
- Sharma, G. R., V. D. Misra, D. Mandal, et al. 1980. *Beginnings of agriculture*. Allahabad, India.
- Shen, J. H. 1980. Rice breeding in China. In *Rice improvement in China and other Asian countries*, 9-36. Los Baños, Philippines.
- Spencer, J. E. 1963. The migration of rice from mainland Southeast Asia into Indonesia. In *Plants and the migration of Pacific peoples*, ed. J. Barrau, 83-9. Honolulu.
- Ting, Y. 1949. Chronological studies of the cultivation and the distribution of rice varieties *keng* and *sen* (in Chinese with English summary). *Sun Yatsen University Agronomy Bulletin* 6: 1-32.
- Ting, Y., ed. 1961. *Chinese culture of lowland rice* (in Chinese). Peking.
- Vishnu-Mittre. 1976. Discussion. In *Early history of agriculture*, Philosophical Transactions of Royal Society of London B275: 141.
- Watabe, T. 1973. Alteration of cultivated rice in Indochina. *Japan Agricultural Research Quarterly* 7.

1985. Origin and dispersal of rice in Asia. *East Asian Cultural Studies* 24: 33-9.
- WRI (World Resources Institute). 1986. *Basic books*. New York.
- Yan, W. M. 1989. Further comments on the origin of rice agriculture in China (in Chinese). *Agricultural Archaeology* 2: 72-83.
- Yanagita, K., H. Ando, T. Morinaga, et al. 1969. *Japanese history of rice plant* (in Japanese). Tokyo.
- Yen, D. E. 1977. Hoabinhian horticulture? The evidence and the questions from northwest Thailand. In *Sunda and Sabul*, ed. J. Allen, J. Golson, and R. Jones, 567-99. London.
- You, X. L. 1976. Several views on the rice grains and bone spades excavated from the fourth cultural level of Homo-tu site (in Chinese). *Wen Wu* 8: 20-3.
1982. A historical study of the genetic resources of rice varieties of our country, II (in Chinese). *Agricultural Archaeology* 1: 32-41.

II.A.8 Rye

Rye As a Grass

Rye (*Secale cereale* L.) is closely related to the genus *Triticum* (which includes bread wheat, durum wheat, spelt, and the like) and has sometimes been included within that genus (Mansfeld 1986: 1447). In fact, it was possible to breed Triticale, a hybrid of *Triticum* and *Secale*, which is cultivated today (Mansfeld 1986: 1449).

Cultivated rye (*Secale cereale*) is also so closely related genetically to the wild rye (*Secale montanum*) that both species would appear to have had the same ancestors. Yet to say that the cultivated rye plant derived from the wild one is an oversimplification because both plants have been changing their genetic makeup since speciation between the wild and cultivated plants first occurred.

The cultigen *Secale cereale* was brought to many parts of the world, but wild rye still grows in the area where cultivated rye originated, which embraces the mountains of Turkey, northwestern Iran, Caucasia, and Transcaucasia (Zohary and Hopf 1988: 64-5; Behre 1992: 142).

The distribution area of wild rye is slightly different from the area of origin of other Near Eastern crops. Wild rye is indigenous to areas north of the range of the wild *Triticum* and *Hordeum* species; these areas have a more continental climate with dry summers and very cold, dry winters. The environmental requirements of cultivated rye reflect these conditions of coldness and dryness: It has a germination temperature of only 1 to 2 degrees Centigrade, which is lower than that of other crops. Indeed, low temperatures are necessary to trigger sprouting (Behre 1992: 145), and the plant grows even in winter if the tem-

perature exceeds 0 degrees Centigrade, although rye can suffer from a long-lasting snow cover. In spring it grows quickly, so that the green plant with unripe grains reaches full height before the summer drought begins (Hegi 1935: 498-9). Obviously, these characteristics make rye a good winter crop. It is sown in autumn, grows in winter and spring, and ripens and is harvested in summer - a growth cycle that is well adapted to continental and even less favorable climatic conditions. There is also another cultigen of rye - summer rye - which is grown as a summer crop. But because of a low yield and unreliability, it is rather uncommon today (Hegi 1935: 497).

Clearly, then, the constitution of the wild grass ancestor of cultivated rye is reflected in the cultivated crop. Rye is predominantly grown as a winter crop, on less favorable soils, and under less favorable climatic conditions than wheat.

The Question of Early Cultivation

There is evidence for the ancient cultivation of rye in the Near East dating back to the Neolithic. Gordon Hillman (1975: 70-3; 1978: 157-74; see also Behre 1992: 142) found cultivated rye in aceramic early Neolithic layers of Tell Abu Hureyra in northern Syria and also at Can Hasan III in central Anatolia. Hillman reports that there were entire rachis internodes at these sites, proof that the selective pressures of cultivation were operating, because only a plant with a nonbrittle rachis can be harvested efficiently. It is not clear, however, if rye was actually cultivated at these Neolithic sites or whether the plant only underwent such morphological adaptations while being sown and harvested as a weedy contaminant of other crops.

To this day, rye remains a vigorous weed in Near Eastern wheat and barley fields, and its nonbrittle rachis internodes resemble a cultivated plant in spite of the fact that it is not intentionally sown. It is harvested together with the more desirable wheat and barley as a "maslin crop" (a crop mixture), and, in climatically unfavorable years, the rye yield is often better than the yield of barley or wheat in these fields. Even an examination of the harvested crop may give the false impression that the rye has been deliberately cultivated. It is interesting to note that such "volunteer" rye is called "wheat of Allah" by Anatolian peasants (Zohary and Hopf 1988: 64) because it is assumed that God "sent" a crop in spite of the bad weather conditions that were unfavorable to the sown wheat.

Possibly this process of unintentionally cultivating rye, while intentionally cultivating wheat and barley, also took place in the early Neolithic fields of Tell Abu Hureyra and Can Hasan III that Hillman investigated. So we do not know if rye was deliberately grown as a crop in its own right or if it was only "wheat of Allah." It is the case that Hillman's evidence

for the early cultivation of rye in the Near East contradicts an earlier opinion by Hans Helbaek (1971: 265–78), who assumed that rye derived from central rather than western Asia.

Rye As a Weed

Rye reached Europe at the dawn of the region's Neolithic Revolution, but probably as a weed. Angela M. Kreuz (1990: 64, 163) has discovered rye remains in Bruchenbrücken, near Frankfurt, in central Germany. This site is dated to the earliest phase of the Linearbandkeramik, which is the earliest phase of agriculture in central Europe. Similarly, Ulrike Piening (1982: 241–5) found single rye grains in a Linearbandkeramik settlement at Marbach, near Stuttgart, in southern Germany. But at both sites only single rye grains were found among great amounts of grains of other species. The same is the case with the few other early rye finds in Europe (Piening 1982: 242–4; Behre 1992: 142–3). Thus, the evidence appears to indicate that rye existed during the early phase of agricultural development in Europe as a weed, and an uncommon one at that.

In the Neolithic, however, most grain cultivation took place on fertile loess soils situated in regions where typical winter crop weeds were not present. Such conditions did not favor rye expansion and, consequently, there was little opportunity to compare the durability of rye to that of *Triticum* species, as was the case with the development of the “wheat of Allah” in the maslin crop fields in the Anatolian mountains.

The proportions of rye, however, were greater in some grain assemblages from Bronze Age sites. Many have assumed that rye was cultivated as a Bronze Age crop, especially in eastern central Europe (Körber-Grohne 1987: 44), but the evidence remains scarce and questionable (Behre 1992: 143). Yet spelt (*Triticum spelta*), a grain similar to rye, was commonly grown in this region during the Bronze Age (Körber-Grohne 1987: 74). Because spelt was normally cultivated as a winter crop, spelt grain assemblages from archaeological sites are contaminated with winter crop weed seeds (Küster 1995: 101). Thus, it could be that the beginning of winter crop cultivation favored the expansion of winter rye as a weed in spelt fields. This was probably the case especially in areas less favorable to agriculture that were being cultivated from the Bronze Age forward, as, for example, in some areas of the Carpathians and the Alps, where rye pollen grains have been recorded several times in layers dating to the Bronze Age (Küster 1988: 117). Definitive evidence of an early rye expansion to the Alps, however, awaits more extensive plant macrofossil examination in these marginal agricultural areas high up in the mountains.

Rye As a Secondary Cultivated Crop

Spelt cultivation, possibly as a winter crop, expanded during the Pre-Roman Iron Age to other parts of Europe (Körber-Grohne 1987: 74), as agriculture itself spread to areas with less fertile soils, such as those of sand and gravel in northern central Europe. These soils, as well as the local ecological conditions of humid climate and light snow cover, favor a winter crop plant that grows during mild winter days and in the spring but will not suffer from summer drought on sandy soils.

Pollen (Küster 1988: 117; Behre 1992: 148) and macrofossil evidence (Behre 1992: 143) show that rye became more common during the Pre-Roman Iron Age, perhaps in those winter crop fields on the less favorable soils just described. At this point, rye was still growing as a weed, but because it had the qualities of a cultivated plant under these ecological conditions, rye eventually predominated in fields planted with spelt. This success is typical of secondary plants that are cultivated by chance within stands of other crops.

Karl-Ernst Behre (1992: 143) has compiled a list of the most ancient finds of pure, or possibly pure, rye cultivated during the Iron Age in Europe. This shows concentrations in the eastern Alps, the countries around the Black Sea, and the western and northern marginal areas of Europe.

But rye became more common during the Roman Age, as populations grew, thus increasing the demand for food. During this time, ever greater amounts of lands with less fertile soils were brought under cultivation, and the expansion of winter crop cultivation provided more reliable and greater yields. Abundant *Secale* grains have been discovered on some Roman sites, giving the impression that rye was cultivated as a main crop (Behre 1992: 143–5). It is, however, unlikely that the Romans themselves propagated rye (with which they were unfamiliar) because climate militated against its growth in the Mediterranean region (Behre 1992: 145). Only a few Roman Age sites outside the Roman Imperium have been examined by archaeobotanists so far, but there is clear evidence that rye was grown outside the Empire as a main crop. A detailed study from an area in northern Germany has shown that the shift to rye cultivation took place during the second century A. D. (Behre 1992: 146).

A few hypotheses for the increased importance of rye have been put forward. For one, rye may have been imported from areas outside to sites inside the Imperium (Dickson and Dickson 1988: 121–6), which suggests increased demand, and, in what is not necessarily a contradiction, Behre (1992: 149–50) emphasizes that the expansion of rye during the Roman Age reflects the improvement of harvesting methods beyond the earlier technique of plucking the grain ear by ear. Because all cultivars depend on harvesting

for seed dispersal, such a thorough method would not have favored the expansion of rye. But during the Iron Age and Roman times, harvesting methods grew more sophisticated, and the advent of new mowing equipment made rye's dispersal more likely.

Another hypothesis involves climatic deterioration as an explanation for the expansion of rye. To date, however, there is no clear evidence for climatic change during the Roman Age. Most likely then, by way of summary, the major reasons for the increased importance of rye cultivation were the expansion of agriculture to more marginal fields, the growing importance of winter crops, and changing harvesting methods.

Medieval Rye Cultivation

During the Middle Ages, rye became a very important crop in many parts of Europe. As agriculture was introduced to marginal mountainous landscapes, the cultivation of rye was frequently the best alternative. More important, although the acid, sandy soils in northern and north-central Europe became exhausted from overcropping, the custom developed of enriching them with "plaggen," which was heath, cut down and transported from the heathlands to the farmlands (Behre 1992: 152). Although this caused a further impoverishment of the already relatively infertile heathlands, such a practice made it possible to control the fertility of marginal fields and to grow crops near the settlements. On these soils "eternal rye cultivation" (Behre 1992: 152) became possible, allowing cropping every year.

In other regions where rye replaced spelt, as for example in southern Germany, such a replacement resulted from practical reasons (Rösch, Jacomet, and Karg 1992: 193–231). Because spelt is a hulled crop, the grains must be dehulled after threshing. This is not necessary with rye or wheat, but the latter is very sensitive to diseases caused by primitive storage conditions in damp environments. Thus, because it was easier to store rye than wheat, and easier to process rye than spelt, rye replaced spelt in many places during the period between the Roman Age and the Middle Ages (Rösch et al. 1992: 206–13). In other areas, of course, such as the mountains of the Ardennes in Belgium and northern France, and the area around Lake Constance, spelt has been grown until recent times and was never replaced by rye.

The relative importance of a grain crop in the various areas of Germany can be determined from the language of historical documents. This is because the term *Korn* ("corn") signifies the most important crop over the ages. So it is interesting to find that in regions where rye cultivation predominated during the Middle Ages and early modern times, the term *Korn* is connected with rye, but in others it is associated with spelt or wheat.

Rye crossed the Atlantic to the New World with colonists heading to both the south and the north of North America. In the south, Alexander von Humboldt, who visited Mexico at the turn of the nineteenth century, discovered rye growing "at heights where the cultivation of maize would be attended with no success" (Humboldt 1972: 97). In addition, he reported that the plant was seldom attacked by a disease that in Mexico "frequently destroys the finest wheat harvests when the spring and the beginning of the summer have been very warm and when storms are frequent" (Humboldt 1972: 104).

In the north, where rye was also extensively cultivated in colonial New England, symptoms of ergotism (a disease caused by ingestion of the ergot fungus that infects many grains, but especially rye) are believed to have often been manifested by the population. Such symptoms (especially those of nervous dysfunction), are seen to have been present in the Salem witchcraft affair, in the "Great Awakening," and in epidemics of "throat distemper" (Matossian 1989). Certainly ergotism had a long and deadly history in Europe, beginning before the early Middle Ages. Some 132 epidemics were counted between 591 and 1789, the last occurring in France during the time of the "Great Fear," which just preceded the French Revolution and which some have seen as leading to it (Haller 1993).

In conclusion, although rye has been said to be our "oldest crop," and baking company advertisements call rye bread the traditional bread, as we have seen, this is certainly not the case. Only gradually did this crop, which began as a weed among cultigens, grow to prominence. But it has also traveled as far from its origins as the United States and Canada (Körber-Grohne 1987: 40), where the winters are cold enough to stimulate the germination of the grains – the same stimulus rye plants received in the mountains of the Near East before they spread out into eastern, central, northern, and western Europe.

Hansjörg Küster

Bibliography

- Behre, Karl-Ernst. 1992. The history of rye cultivation in Europe. *Vegetation History and Archaeobotany* 1: 141–56.
- Dickson, C., and J. Dickson. 1988. The diet of the Roman army in deforested central Scotland. *Plants Today* 1: 121–6.
- Haller, John S., Jr. 1993. Ergotism. In *The Cambridge world history of human disease*, ed. Kenneth F. Kiple, 718–19. Cambridge and New York.
- Hegi, Gustav. 1935. *Illustrierte Flora von Mittel-Europa*, Vol. 1. Second edition. Munich.
- Helbaek, Hans. 1971. The origin and migration of rye, *Secale cereale* L.; a palaeo-ethnobotanical study. In *Plant Life*

- of South-West Asia, ed. P. H. Davis, P. C. Harper, and I. G. Hedge, 265–80. Edinburgh.
- Hillman, Gordon. 1975. The plant remains from Tell Abu Hureyra: A preliminary report. In A. M. T. Moore et al., *Excavations at Tell Abu Hureyra in Syria: A preliminary report. Proceedings of the Prehistoric Society* 41: 70–3.
1978. On the origins of domestic rye – *Secale cereale*: The finds from aceramic Can Hasan III in Turkey. *Anatolian Studies* 28: 157–74.
- Humboldt, Alexander von. 1972. *Political essay on the kingdom of New Spain*, ed. Mary M. Dunn. Norman, Okla.
- Körber-Grohne, Udelgard. 1987. *Nutzpflanzen in Deutschland*. Stuttgart.
- Kreuz, Angela M. 1990. Die ersten Bauern Mitteleuropas. Eine archäobotanische Untersuchung zu Umwelt und Landwirtschaft der ältesten Bandkeramik. *Analecta Praehistorica Leidensia* 23. Leiden.
- Küster, Hansjörg. 1988. *Vom Werden einer Kulturlandschaft*. Weinheim.
1995. *Postglaziale Vegetationsgeschichte Südbayerns. Geobotanische Studien zur prähistorischen Landschaftskunde*. Berlin.
- Mansfeld, Rudolf. 1986. *Verzeichnis landwirtschaftlicher und gärtnerischer Kulturpflanzen*, ed. Jürgen Schultze-Motel. Second edition. Berlin.
- Matossian, Mary Kilbourne. 1989. *Poisons of the past: Molds, epidemics, and history*. New Haven, Conn., and London.
- Piening, Ulrike. 1982. Botanische Untersuchungen an verkohlten Pflanzenresten aus Nordwürttemberg. Neolithikum bis Römische Zeit. *Fundberichte aus Baden-Württemberg* 7: 239–71.
- Rösch, Manfred, Stefanie Jacomet, and Sabine Karg. 1992. The history of cereals in the region of the former Duchy of Swabia (Herzogtum Schwaben) from the Roman to the post-medieval period: Results of archaeobotanical research. *Vegetation History and Archaeobotany* 1: 193–231.
- Zohary, Daniel, and Maria Hopf. 1988. *Domestication of plants in the Old World*. Oxford.

II.A.9 🌾 Sorghum

Grain sorghum (*Sorghum bicolor* [Linn.] Moench) is a native African cereal now also widely grown in India, China, and the Americas. Sorghum ranks fifth in world cereal grain production, and fourth in value (after rice, wheat, and maize) as a cereal crop. It is grown on 40 to 50 million hectares annually, from which up to 60 million metric tons of grain are harvested. In Africa and Asia traditional cultivars are grown, usually with low agricultural inputs, and average yields are below 1 metric ton per hectare. But more than 3 metric tons of grain are harvested per acre in the Americas, where farmers plant modern sorghum hybrids. Sorghum is more tolerant to drought and better adapted for cultivation on saline soils than is maize. It holds tremendous promise as a cereal to feed the rapidly expanding populations of Africa and Asia. In the Americas it is replacing maize as an animal feed.

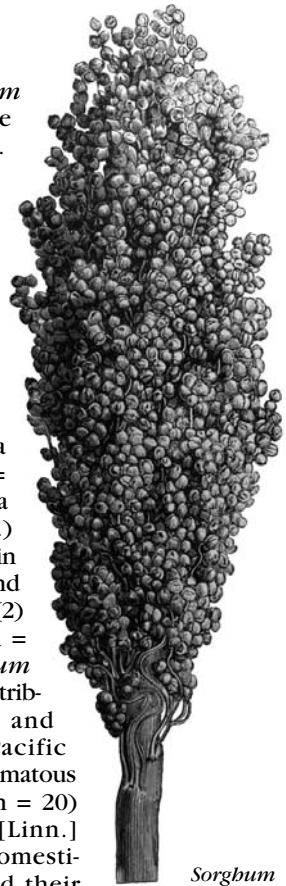
Morphology and Distribution

The grass genus *Sorghum* Moench is one of immense morphological variation. It is taxonomically subdivided into sections *Chaetosorghum*, *Heterosorghum*, *Parasorghum*, *Stiposorghum*, and *Sorghum* (Garber 1950), and these sections are recognized as separate genera by W. D. Clayton (1972). The genus *Sorghum* is here recognized to include: (1) a complex of tetraploid ($2n = 40$) rhizomatous taxa (*S. halapense* [Linn.] Pers.) that are widely distributed in the Mediterranean region and extend into tropical India; (2) a rhizomatous diploid ($2n = 20$) species (*S. propinquum* [Kunth] Hitchc.) that is distributed in Southeast Asia and extends into adjacent Pacific Islands; and (3) a nonrhizomatous tropical African diploid ($2n = 20$) complex (*S. bicolor* [Linn.] Moench) that includes domesticated grain sorghums and their closest wild and weedy relatives

(de Wet and Harlan 1972). Genetic introgression is common where wild rhizomatous or spontaneous nonrhizomatous taxa become sympatric with grain sorghums, and derivatives of such introgression have become widely distributed as weeds in sorghum-growing regions.

The domesticated sorghum complex is morphologically variable. It includes wild, weed, and domesticated taxa that are divided by J. D. Snowden (1936, 1955) among 28 cultivated species, 13 wild species, and 7 weed species. Following the classification of cultivated plants proposed by Jack R. Harlan and J. M. J. de Wet (1972), the wild taxa are recognized as subspecies *verticilliflorum* (Steud.) de Wet, the weed taxa as subspecies *drummondii* (Steud.) de Wet, and the grain sorghums as subspecies *bicolor* (de Wet and Harlan 1978).

Subspecies *verticilliflorum* includes races *verticilliflorum*, *arundinaceum*, *virgatum*, and *aethiopicum*. These grade morphologically and ecologically so completely into one another that they do not deserve formal taxonomic rank. This subspecies is indigenous to tropical Africa but has become widely distributed as a weed in tropical Australia (de Wet, Harlan, and Price 1970). It differs from grain



Sorghum

sorghum primarily in being spontaneous rather than cultivated and in being capable of natural seed dispersal.

Verticilliflorum is the most widely distributed, and morphologically the most variable, race of the subspecies. It extends naturally across the African savannah, from Senegal to the Sudan and South Africa. It is distinguished from the other races by its large and open inflorescences with long and spreading branches. *Verticilliflorum* is an aggressive colonizer of naturally disturbed habitats, and it often forms large continuous populations in flood plains. It is commonly harvested as a wild cereal in times of scarcity.

Race *arundinaceum* is distributed along the margins of tropical forests of the Congo basin. It is sympatric with *verticilliflorum* along the transition zone between savannah and forest, and the races introgress. Derivatives of such hybridization aggressively colonize areas of forest that are cleared for agriculture. *Arundinaceum* is typically characterized by large and open inflorescences with long branches that become pendulous at maturity.

Race *virgatum* occurs along stream banks and irrigation ditches in arid regions of tropical northeastern Africa. Wild populations are harvested as a cereal during times of famine. It is widely sympatric with race *verticilliflorum*, and gene exchange between them is common. It typically has smaller inflorescences than *verticilliflorum*.

Race *aethiopicum* is drought tolerant. It extends across the West African Sahel and into the Sudan. In flood plains, it frequently forms large continuous populations and is harvested as a wild cereal. The distribution and habitat of *aethiopicum* rarely overlap with the other races. It is characterized by large spikelets that are densely tomentose.

Subspecies *drummondii* is an obligate weed derived through introgression between subspecies *verticilliflorum* and cultivated grain sorghums. It became widely distributed across tropical Africa as part of cereal agriculture. Morphological variation is extensive as a result of hybridization among the different races of grain sorghum and different races of close wild relatives. Stabilized derivatives of such introgression accompanied the cereal to India and the highlands of Ethiopia. Weeds often resemble grain sorghums in spikelet morphology, but they retain the ability of natural seed dispersal.

Grain sorghums also introgress with the Eurasian *S. halepense* to form diploid or tetraploid weedy derivatives. Johnson grass of the American Southwest and some sorghums of Argentina are tetraploid derivatives of such introgression. Diploid derivatives of hybridization between grain sorghum and Johnson grass have recently become obnoxious weeds in the American corn belt.

Subspecies *bicolor* includes all domesticated grain sorghums. The 28 cultivated species recognized by Snowden (1936) are artifacts of sorghum cultivation.

They represent selections by farmers for specific adaptations and food uses, and they do not deserve formal taxonomic rank. Grain sorghums are classified by Harlan and de Wet (1972) into races *bicolor*, *kafir*, *caudatum*, *durra*, and *guinea*. Sorghums belonging to different races hybridize where they are grown sympatrically, and cultivars have become established that combine characteristics of two or more of these races. Extensive racial evolution took place in Africa before sorghum was introduced as a cereal into Asia (Harlan and Stemler 1976).

Race *bicolor* resembles spontaneous weedy sorghums in spikelet morphology, but all cultivars depend on harvesting for seed dispersal. Mississippi chicken corn probably represents a derivative of abandoned cultivated race *bicolor* that entered America during the slave trade. It is spontaneous and must have regained the ability of natural seed dispersal through mutation. *Bicolor* sorghums are characterized by open inflorescences, having spikelets with long and clasping glumes that enclose the grain at maturity. Some cultivars of race *bicolor* are relics of the oldest domesticated sorghums, whereas others are more recent derivatives of introgression between evolutionally advanced cultivars and spontaneous sorghums.

Bicolor sorghums are widely distributed in Africa and Asia but are rarely of major economic importance because of their low yield. Cultivars survive because they were selected for specific uses. They are grown for their sweet stems (chewed as a delicacy), for the high tannin content of the grains (used to flavor sorghum beer), and for use as fodder. Cultivars often tiller profusely, which tends to make their sweet stems desirable as fodder for livestock in Africa.

Race *kafir* is the most common cultivated sorghum south of the equator in Africa. It never became widely distributed in India and China, probably because of limited trade between southern Africa and India or the Near East before colonial times. Race *kafir* is characterized by compact inflorescences that are cylindrical in shape. Spikelets have glumes that tightly clasp the usually much longer mature grain. Sorghum has been replaced by maize in areas with high rainfall, but *kafir* sorghums remain the most important cereal crop of the southern savannahs in areas with between 600 and 900 millimeters (mm) of annual rainfall. At the drier limits of agriculture, sorghum is replaced as a cereal by pearl millet. In the wettest parts, sorghum competes as a food cereal not only with maize but also with finger millet. The grain of *kafir* sorghums is commonly high in tannin. This provides partial protection against bird damage and confers resistance to grain molds that reduce grain quality. Tannin also, however, reduces the digestibility of porridges produced from *kafir* sorghums, which today are grown mainly to produce malt for the making of a highly nutritious beer. This beer is commercially produced in Zimbabwe and South Africa.

Race *caudatum* is distinguished by its asymmetrical grains. The grain is usually exposed between the glumes at maturity, with the embryo side bulging and the opposite side flat or concave. Inflorescences range from very compact to rather open with spreading branches. *Caudatum* cultivars are highly adaptive and are grown in areas with as low as 350 mm and as high as 1,000 mm of annual rainfall. Selected cultivars are resistant to fungal leaf diseases, to ergot of the grain, and to infestation by insects or the parasitic striga weed. *Caudatum* sorghums are a major food source of people speaking Chari-Nile languages in the Sudan, Chad, Uganda, northeastern Nigeria, and Cameroon (Stemler, Harlan, and de Wet 1975). Along the flood plains of the Niger River in Chad, *caudatum* sorghums are grown in nurseries and transplanted to cultivated fields as flood waters recede (Harlan and Pasguereau 1969). The grains are ground into flour from which a fermented porridge is produced. *Caudatum* sorghums are also commercially grown in Nigeria for the production of malt used in the brewing industry.

Race *durra* is the most drought tolerant of grain sorghums. Selected cultivars mature in less than three months from planting, allowing escape from terminal drought stress in areas with short rainy seasons. The name *durra* refers to the Arabic word for sorghum, and the distribution of *durra* sorghums in Africa is closely associated with the spread of Islam across the Sahel. The grain is also extensively grown in the Near East, China, and India. Inflorescences are usually compact. Spikelets are characteristically flattened and ovate in outline, with the lower glume either creased near the middle or having a tip that is distinctly different in texture from the lower two-thirds of the glume. Grains are cooked whole after decortication or are ground into flour to be prepared as porridge or baked into unleavened bread.

Race *guinea* is distinguished by long glumes that tightly clasp the obliquely twisted grain, which becomes exposed between them at maturity. Inflorescences are large and often open, with branches that become pendulous at maturity. These are adaptations for cultivation in areas with high rainfall, and *guinea* is the principal sorghum of the West African *Guinea* zone with more than 800 mm of annual rainfall. *Guinea* sorghums are also grown along the high-rainfall highlands from Malawi to Swaziland and in the ghats of Central India. It is a principal food grain in West Africa and Malawi. In Senegal, the small and hard grains of an indigenous cultivar are boiled and eaten, similar to the way rice is prepared and consumed in other parts of the world. In Malawi, the sweet grains of a local cultivar are eaten as a snack while still immature. *Guinea* sorghums are valued for the white flour that is produced from their tannin-free grains.

Intermediate races recognized by Harlan and de Wet (1972) include sorghum cultivars that are not readily classifiable into any one of the five basic races. They combine characteristics of race *bicolor* with

those of the other four basic races, of *guinea* and *caudatum*, or of *guinea* and *kafir*. Cultivars with intermediate morphologies occur wherever members of basic races are grown sympatrically in Africa. Intermediate cultivars have become widely distributed in India. Modern high-yielding sorghum hybrids combine traits of races *kafir*, *durra*, and *caudatum* in various combinations.

Domestication and Evolutionary History

Cereal domestication is a process, not an event. Domestication is initiated when seeds from planted populations are harvested and sown in human-disturbed habitats (in contrast to naturally disturbed habitats), and it continues as long as the planting and harvesting processes are repeated in successive generations (Harlan, de Wet, and Price 1973). The initial ability to survive in disturbed habitats is inherent in all wild grasses that were adopted as cereals. In fact, as aggressive colonizers, they can form large continuous colonies in naturally disturbed habitats. This weedy characteristic of these plants facilitates harvesting and eventually leads to their domestication. Sowing in cultivated fields reinforces adaptation for survival in disturbed habitats, and harvesting of sown populations selects against mechanisms that facilitate natural seed dispersal. Thus, domesticated cereals have lost the ability to compete successfully for natural habitats with their wild relatives. They depend on farming for suitable habitats and on harvesting and sowing for seed dispersal.

There is little doubt that subspecies *verticilliflorum* gave rise to grain sorghums under domestication. This spontaneous complex of tropical African sorghums is an aggressive colonizer of naturally disturbed habitats, and because it forms large continuous stands, it remains a favorite wild cereal of nomads as well as farmers during times of food scarcity. Snowden (1936) and R. Porteres (1962) have suggested that race *arundinaceum* (of forest margins) gave rise to *guinea* sorghums, the desert race *aethiopicum* to *durra* sorghums, and the savannah race *verticilliflorum* to *kafir* sorghums. Distribution and ethnological isolation certainly suggest three independent domestications of grain sorghum. This, however, is unlikely. Close genetic affinities between specific cultivated races and the spontaneous races with which they are sympatric resulted from introgression. Such introgression continues between grain sorghums and their close, spontaneous relatives. Racial evolution of advanced cultivated races resulted from selection by farmers who grew bicolor sorghums for specific uses, and from natural adaptations to local agro-ecological environments.

The wild progenitor of cultivated sorghums is the widely distributed race *verticilliflorum*. It could have been domesticated anywhere across the African savanna. Jack Harlan (1971) proposes that sorghum

was taken into cultivation along a broad band of the savanna from the Sudan to Nigeria, where verticilliflorum is particularly abundant. H. Dogget (1965) previously had suggested that the initial domestication occurred in the northeastern quadrant of Africa, probably undertaken by early farmers in Ethiopia who learned from the ancient Egyptians how to grow barley and wheat. These two Near Eastern cereals have been grown in Egypt and along the Mediterranean coast of North Africa since at least the fifth century B.C. (Clark 1971).

Tropical agriculture in Africa must have started in the savannah along the southern fringes of the Sahara (Clark 1976, 1984). Archaeological evidence indicates that pearl millet (*Pennisetum glaucum* [Linn.] R. Br.), sorghum, and finger millet (*Eleusine coracana* [Linn.] Gaertn.) were among the earliest native cereals of the savannah to be domesticated.

J. S. Wigboldus (1991) has suggested that there is little evidence to indicate cereal cultivation south of the Sahara before the ninth century of the Christian era. Archaeological evidence, however, indicates that cereal agriculture in the African savanna is much older than this. Inhabitants of the Dhar Tichitt region of Mauritania, extending from the middle of the second to the middle of the first millennium B.C., evidently experimented with the cultivation of native grasses (Munson 1970). During the first phase of settlement, bur grass (*Cenchrus biflorus* Roxb.) seems to have been the most common grass harvested as a wild cereal. It is still extensively harvested in the wild as a source of food during times of scarcity. In the middle phases, *Brachiaria deflexa* (Shumach.) Hubbard, now cultivated on the highlands of Mali (Porteres 1976), and *Pennisetum glaucum*, now grown as pearl millet across the Sahel, became equally common, as shown by their impressions on potsherds. In later phases, starting about 1000 B.C., impressions of what is almost certainly domesticated pearl millet became dominant (Munson 1970). It is not possible, however, to determine whether pearl millet was domesticated at Dhar Tichitt or whether this cereal was introduced to these settlements from other parts of the West African Sahel.

Tropical African grasses were also grown as cereals in eastern Africa before the beginning of the Christian era. Potsherds from a Neolithic settlement at Kadero in the central Sudan, dated to between 5,030 and 5,280 years ago, reveal clear impressions of domesticated sorghum and finger millet spikelets and grains (Klichowska 1984). Both cereals are today extensively grown in eastern and southern Africa. Indirect evidence of early sorghum and finger millet cultivation in Africa comes from the presence of these African cereals in Neolithic settlements of India, dated to about 1000 B.C. (Weber 1991). Other archaeological evidence indicates that sorghum cultivation spread from eastern Africa to reach northeastern Nigeria not later than the tenth century A.D. (Connah 1967) and, together with

pearl millet and finger millet, reached southern Africa not later than the eighth century A.D. (Shaw 1976).

That native grasses were grown as cereals not less than 3,000 years ago along the southern fringes of the Sahara is not surprising. Wheat and barley were grown in Egypt and along the Mediterranean coast of North Africa by the latter part of the fifth millennium B.C. (Shaw 1976), and the knowledge of cereal agriculture reached the highlands of Ethiopia some 5,000 years ago. These Near Eastern cereals cannot be grown successfully as rain-fed crops in lowland tropics. Experimentation with the cultivation of native grasses in the semiarid tropical lowlands seems a logical next step in the development of African plant husbandry. Nor is the absence of domesticated sorghum in West Africa before the tenth century A.D. surprising. Sorghum is poorly adapted to the arid Sahel, where finger millet was domesticated and remains the principal cereal, and an abundance of wild food plants and animals probably made agriculture in the Guinea zone less productive than hunting and gathering during the beginnings of plant husbandry in tropical Africa. Racial evolution gave rise to races guinea, caudatum, durra, and kafir and is associated with adaptation to agro-ecological zones and the isolation of different ethnic groups who adopted sorghum cultivation. Morphological differentiation took place in Africa, except for race durra that may have evolved in Asia after sorghum cultivation became established in southwestern Asia.

Race guinea's open panicles and spikelets with widely gaping glumes are adaptations for successful cultivation in high-rainfall areas. The glumes enclose the immature grain to protect it from infection by grain molds, but they gape widely at maturity to allow the grain to dry rapidly after a rain and thus escape damage. Guinea sorghums, which probably evolved in Ethiopia, are still grown in the Konso region, and from there they may have spread along the mountains south to Swaziland and west to the Guinea coast. Cultivated sorghum belonging to race guinea was already growing in Malawi during the ninth century A.D. (Robbinson 1966). Today, almost half the sorghum production in Nigeria comes from guinea sorghums.

Kafir sorghums evolved south of the equator and never became widely distributed outside the southern African savanna. They are probably relatively recent in origin. Kafir sorghums became associated with Iron Age Bantu settlements only during the eighth century A.D. (Fagan 1967; Phillipson and Fagan 1969). Kafir sorghums are genetically more closely allied to local verticilliflorums than to other spontaneous sorghums. This led Y. Schechter and de Wet (1975) to support Snowden's (1936) conclusion that race kafir was independently domesticated from other sorghums in southern Africa. It is more likely, however, that kafir sorghums were derived from introduced bicolor sorghums that introgressed with local wild sorghum adapted to the arid southern savanna.

As already mentioned, race durra is the most drought-tolerant of all grain sorghums. Their wide distribution in semiarid Asia caused Harlan and A. B. L. Stemler (1976) to propose that durra sorghums evolved in West Asia from earlier introductions of race bicolor. Archaeological remains indicate that bicolor sorghums were grown in India not later than the early first millennium B.C. (Weber 1991). Durras remain the common cultivated sorghums in semiarid Asia. In Africa, they are grown across the Sahel, and their distribution seems to be associated with the expansion of Islam across North Africa.

The cultivation of caudatum sorghums is closely associated in Africa with the distribution of people who speak Chari-Nile languages (Stemler, Harlan, and de Wet 1975). Caudatum sorghums probably represent selections from race bicolor in the eastern savannah during relatively recent times. Bicolor sorghums were important in the Sudan as late as the third century A.D., and archaeological sorghum remains from Qasr Ibrim and Jebel et Tomat in the Sudan belong to race bicolor (Clark and Stemler 1975). The beautifully preserved sorghum inflorescences from Qasr Ibrim date from the second century (Plumley 1970). The only known archaeological remains of caudatum are those from Daima, dated A.D. 900 (Connah 1967). Introgression of caudatum with durra sorghums of the Sahel and with guinea sorghums of West Africa gave rise to a widely adapted complex that is extensively used in modern sorghum breeding.

The spread of sorghum as a cereal to Asia is poorly documented. Carved reliefs from the palace of Sennacherib at Nineveh are often cited as depicting cultivated sorghum (see Hall 1928, plates 30 and 32). But these plants were actually the common reed (*Phragmites communis* Trin.) growing along the edges of a marsh with pigs grazing among them, certainly not a habitat for growing sorghum. Similar plants appear in imperial Sassanian hunting scenes from Iran (for illustrations, see Reed 1965).

In the Near East, sorghum is an important cereal only in Yemen. Sorghum probably reached India directly from East Africa during the latter part of the second century B.C. (Vishnu-Mittre and Savithri 1982), and in India, race durra evolved. From India durra sorghum was introduced to China, probably during the Mongol conquest (Hagerthy 1940), and to the Sahel during the expansion of Islam across northern Africa. Introduction into the New World most likely started with the slave trade between Africa and the Americas. The weedy Mississippi chicken corn may represent an escape from cultivation dating back to colonial times.

Sorghum As a World Cereal

Sorghum is an important rain-fed cereal in the semiarid tropics. Production in recent years has been between 50 and 60 million metric tons of grain har-

vested from around 45 million hectares. The major production areas are North America (excluding Mexico) with 34 percent of total world production, Asia (32 percent), Africa (26 percent), and South America (6 percent). The Caribbean, Meso America, and South America together account for about 17 percent of world sorghum production, with Mexico producing almost 59 percent of this amount. Potential yield of improved sorghum hybrids under rain-fed agricultural conditions is well over 6 metric tons per hectare. Actual maximum yields are closer to 4 metric tons, and average yields are about 1.5 metric tons per hectare. Sorghum is often grown on marginal agricultural land. In Africa and Asia, where local cultivars are still extensively grown with a minimum of agricultural inputs, average yield is well below 1 metric ton per hectare. Sorghum is grown as a cereal for human consumption in Africa and Asia and as animal feed in the Americas and Australia. Sorghum is also extensively grown as a fodder crop in India.

Sorghum production in Africa extends across the savanna in areas with as little as 300 mm and as much as 1,500 mm of annual rainfall. At the drier limits of its range, sorghum is replaced in Africa by pearl millet, in India by pearl millet or foxtail millet (*Setaria italica* [Linn.] P. Beauv.), and in China by foxtail millet. In areas with more than 900 mm of annual rainfall, maize has replaced sorghum across tropical Africa since its introduction from America during the sixteenth century.

Major factors limiting yield in Africa are infestation of cultivated fields by *Striga* (a parasitic weed) and the abundance of birds that feed on sorghum grain before it is ready for harvest. Some degree of resistance to bird damage is conferred by high tannin content in developing grains. Tannin, unfortunately, reduces the desirability of sorghum as a cereal grain. Digestibility is improved through fermentation, and fermented food products produced from sorghum grain are extensively used where high tannin cultivars are grown.

Striga is parasitic on most cereals and several broad-leaved crops grown in Africa and India. It produces large numbers of seeds and can become so abundant that fields eventually have to be abandoned. Control of *Striga* requires high agricultural inputs, the most important of which is high soil fertility and weeding. Neither is affordable under conditions of subsistence farming. Some local sorghum cultivars are resistant to *Striga*, but these have low grain yield. Attempts to transfer genes for resistance into more desirable genotypes of sorghum are high in priority for breeding projects in West and East Africa, where *Striga* has become a particularly obnoxious weed.

In Asia, the major sorghum-producing countries are China, India, Thailand, Pakistan, and Yemen. In Thailand, sorghum is grown as a dry-season crop, after a rain-fed crop, usually maize, has been harvested. In India, sorghum is grown as a rain-fed crop (kharif) or

a dry-season crop (rabi), usually following rice or cotton on soils with good moisture retention. Kharif sorghum is usually mixed with pigeon pea in the field. Sorghum matures and is harvested after 90 to 120 days, allowing the season-long-developing pigeon pea an opportunity to mature without competition.

Kharif sorghums were selected for their ability to mature before the end of the rainy season in order to escape terminal drought stress that severely reduces yield. These cultivars are highly susceptible to infection by grain molds, which greatly reduces the desirability of kharif sorghum as a cereal grain. Market samples have revealed that in central and southern India as much as 70 percent of food sorghum grown during the rainy season is infected with grain molds. Cultivars with high tannin content in developing grains are resistant to infection by grain molds, but their flour yields a poor-quality unleavened bread, the major product of sorghum preparation as a food in India. Long-term breeding programs to produce grain-mold-resistant sorghums with grain acceptable to consumers have consistently failed.

Rabi sorghums escape infection by grain molds as they are grown in the dry season, but yields are low because of terminal drought stress. Prices in the market for these sorghums, however, are sufficiently attractive to make rabi sorghum a major crop in India. Production is well below demand, and attempts to shorten the growing season of rabi sorghums to escape drought and at least maintain yield potential are showing promise. Terminal drought stress commonly leads to lodging of these sorghums because of a combination of infection by stem rot fungi and plant senescence. Lodging makes harvesting difficult and contributes to reduced grain quality. To improve stalk quality and overcome lodging, plant breeders in India introduced genes for delayed senescence into high-yielding cultivars. This allows grain harvest when the stalk is still juicy and the leaves are green. Delayed senescence also greatly improves fodder quality. The stalks of both kharif and rabi sorghums are in demand as animal feed. Around urban areas, the demand by the dairy industry for fodder far exceeds the supply, and farmers often derive a higher income from sorghum stalks than sorghum grain.

Shortage of sorghum grain as a food largely excludes its use as animal feed in Africa and Asia. The grain is eaten in a variety of preparations that vary within and between regions. Grains are ground into flour from which unleavened bread is baked, or the flour is used to produce both fermented and unfermented porridges. The grains are also cracked or decorticated and boiled like rice, or whole grains are popped in heated oil.

Commercial grain sorghum production in Africa and Asia is determined by the availability of reliable supplies of the much-preferred rice, wheat, or maize. Only where these three cereals are not available at competitive prices is sorghum an important commer-

cial crop. In China, sorghum is commercially grown for the production of a popular alcoholic beverage. It is used as a substitute for barley malt in the Nigerian beer industry. In southern Africa, a highly nutritious, low-alcohol beer is commercially produced from sorghum malt and flour, and in Kenya sorghum is used to produce a widely accepted baby food. However, attempts in many countries to replace wheat flour partially with sorghum flour in the baking industry have, so far, met with limited success, even though the quality of the bread is acceptable.

In the Americas, sorghum production is determined by demand for the grain as an animal feed. World feed use of sorghum has reached 40 million metric tons annually, with the United States, Mexico, and Japan the main consumers (Food and Agriculture Organization 1988). These three countries used almost 80 percent of the world's sorghum production in 1993. Although North American demand for sorghum grain has stabilized, in South America, where more than 1 million hectares are under sorghum cultivation, demand exceeds production by about 10 percent annually. This shortfall, predicted to increase throughout the next decade, is now mostly made up by imports from the United States. In the quest for self-sufficiency in animal feed, sorghum cultivation in South America is expanding into areas too dry for successful production of maize and into the seasonally flooded Llanos (with acid soils), where sorghum is more productive than maize.

In Asia, the area under sorghum cultivation is declining to make room for the production of fruits, vegetables, and other foods needed to supply rapidly increasing urban populations. Grain production, however, has remained essentially stable in Asia during the last decade because farmers increasingly grow improved cultivars associated with improved farming practices. This allows production to keep pace with demand, except during drought years when the demand for sorghum as human food far exceeds production.

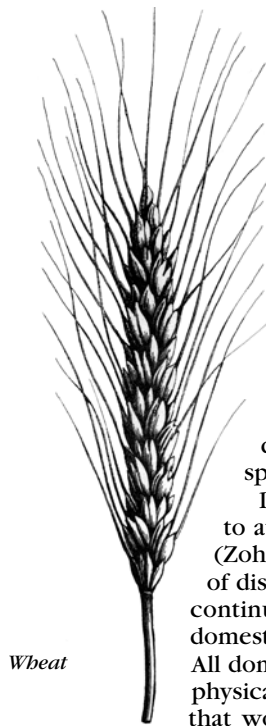
In several African countries, population increase exceeds annual increase in food production. The Food and Agriculture Organization of the United Nations predicted that 29 countries south of the Sahara would not be able to feed their people as the twenty-first century opened. The concomitant increase in demand for cereals will have to be met by the expansion of production into marginal agricultural land, the growing of improved cultivars, and improved farming practices. Pearl millet is the cereal of necessity in areas with between 300 and 600 mm of annual rainfall, and sorghum is the most successful cereal to grow in areas with between 600 and 900 mm of annual rainfall. Because of that, the future of sorghum as a food cereal in Africa and Asia, and as a feed grain in the Americas and Australia, seems secure.

Bibliography

- Clark, J. D. 1971. Evidence for agricultural origins in the Nile Valley. *Proceedings of the Prehistory Society* 37: 34-79.
1976. Prehistoric populations and pressures favoring plant domestication in Africa. In *Origins of African plant domestication*, ed. J. R. Harlan, J. M. J. de Wet, and A. B. L. Stemler, 67-105. The Hague.
1984. Epilogue. In *Origins and early development of food-producing cultures in north-eastern Africa*, ed. L. Krzyzaniak and M. Kobusiewicz, 497-503. Poznan, Poland.
- Clark, J. D., and A. B. L. Stemler. 1975. Early domesticated sorghum from Sudan. *Nature* 254: 588-91.
- Clayton, W. D. 1972. The awned genera of Andropogoneae. *Kew Bulletin* 27: 457-74.
- Connah, G. 1967. Progress report on archaeological work in Bornu in 1964-1966. *Northern history research scheme, Second Interim Report, Zaria*: 20-31.
- de Wet, J. M. J., and J. R. Harlan. 1972. The origin and domestication of *Sorghum bicolor*. *Economic Botany* 25: 128-5.
1978. Systematics and evolution of *Sorghum* sect. *Sorghum* (Gramineae). *American Journal of Botany* 65: 477-84.
- de Wet, J. M. J., J. R. Harlan, and E. G. Price. 1970. Origin of variability in the spontanea complex of *Sorghum bicolor*. *American Journal of Botany* 57: 704-7.
- Dogget, H. 1965. The development of cultivated sorghums. In *Crop plant evolution*, ed. Joseph Hutchinson, 50-69. London.
- Fagan, B. M. 1967. *Iron age cultures in Zambia*. London.
- Food and Agriculture Organization of the United Nations. 1988. *Structure and characteristics of the world sorghum economy*. Committee on Commodity Problems, International Group on Grains, Twenty-third Session. Rome.
- Garber, E. D. 1950. Cytotaxonomic studies in the genus *Sorghum*. *University of California Publications in Botany* 23: 283-361.
- Hagerthy, M. 1940. Comments on writings concerning Chinese sorghums. *Harvard Journal of Asiatic Studies* 5: 234-60.
- Hall, H. R. 1928. *Babylonian and Assyrian sculpture in the British museum*. Paris.
- Harlan, J. R. 1971. Agricultural origins: Centers and non-centers. *Science* 174: 468-74.
- Harlan, J. R., and J. M. J. de Wet. 1972. A simplified classification of cultivated sorghum. *Crop Sciences* 12: 172-6.
- Harlan, J. R., J. M. J. de Wet, and E. G. Price. 1973. Comparative evolution of cereals. *Evolution* 27: 311-25.
- Harlan, J. R., and J. Pasguereau. 1969. Decrue agriculture in Mali. *Economic Botany* 23: 70-4.
- Harlan, J. R., and A. B. L. Stemler. 1976. Races of sorghum in Africa. In *Origins of African plant domestication*, ed. J. R. Harlan, J. M. J. de Wet, and A. B. L. Stemler, 466-78. The Hague.
- Klichowska, M. 1984. Plants of the neolithic Kadero (central Sudan): A palaeoethnobotanical study of the plant impressions on pottery. In *Origin and early development of food-producing cultures in north-eastern Africa*, ed. L. Krzyzaniak and M. Kobusiewicz, 321-6. Poznan, Poland.
- Munson, P. J. 1970. Correction and additional comments concerning the "Tichitt Tradition." *West African Archaeological Newsletter* 12: 47-8.
- Phillipson, D. W., and B. Fagan. 1969. The date of the Ingombe Ilede burials. *Journal of African History* 10: 199-204.
- Plumley, J. M. 1970. Quasar Ibrim 1969. *Journal of Egyptian Archaeology* 56: 12-18.
- Porteres, R. 1962. Berceaux agricoles primaires sur le continent africain. *Journal of African History* 3: 195-210.
1976. African cereals: Eleusine, fonio, black fonio, teff, brachiaria, paspalum, pennisetum, and African rice. In *Origins of plant domestication*, ed. J. R. Harlan, J. M. J. de Wet, and A. B. L. Stemler, 409-52. The Hague.
- Reed, C. A. 1965. Imperial Sassanian hunting of pig and fallow-deer, and problems of survival of these animals today in Iran. *Postillia* 92: 1-23.
- Robbinson, K. R. 1966. The Leopard's kopje culture, its position in the Iron Age of southern Rhodesia. *South African Archaeological Bulletin* 21: 5-51.
- Schechter, Y., and J. M. J. de Wet. 1975. Comparative electrophoresis and isozyme analysis of seed proteins from cultivated races of sorghum. *American Journal of Botany* 62: 254-61.
- Shaw, T. 1976. Early crops in Africa: A review of the evidence. In *Origins of African plant domestication*, ed. J. R. Harlan, J. M. J. de Wet, and A. B. L. Stemler, 107-53. The Hague.
- Snowden, J. D. 1936. *The cultivated races of Sorghum*. London.
1955. The wild fodder sorghums of section *Eu-Sorghum*. *Journal of the Linnean Society: Section Botany* 55: 191-260.
- Stemler, A. B. L., J. R. Harlan, and J. M. J. de Wet. 1975. Caudatum sorghums and speakers of Chari-Nile languages in Africa. *Journal of African History* 16: 161-83.
- Vishnu-Mittre and R. Savithri. 1982. Food economy of the Harappan. In *Harappan civilization*, ed. G. L. Possehl, 205-22. New Delhi.
- Weber, S. A. 1991. *Plants and Harappan subsistence*. New Delhi.
- Wigboldus, J. S. 1991. Pearl millet outside northeast Africa, particularly in northern West Africa: Continuously cultivated from c. 1350 only. In *Origins and development of agriculture in East Africa: The ethnosystems approach to the study of early food production in Kenya*, ed. R. E. Leaky and L. J. Slikkerveer, 161-81. Ames, Iowa.

II.A.10 Wheat

Wheat, a grass that today feeds 35 percent of the earth's population, appeared as a crop among the world's first farmers 10,000 years ago. It increased in importance from its initial role as a major food for Mediterranean peoples in the Old World to become the world's largest cereal crop, feeding more than a billion people in the late twentieth century (Feldman 1976: 121). It spread from the Near East, where it first emerged in the nitrogen-poor soils of a semi-arid Mediterranean climate, to flourish in a wide range of environments - from the short summers of far northern latitudes, to cool uplands, to irrigated regions of the tropics. The real story of its origins



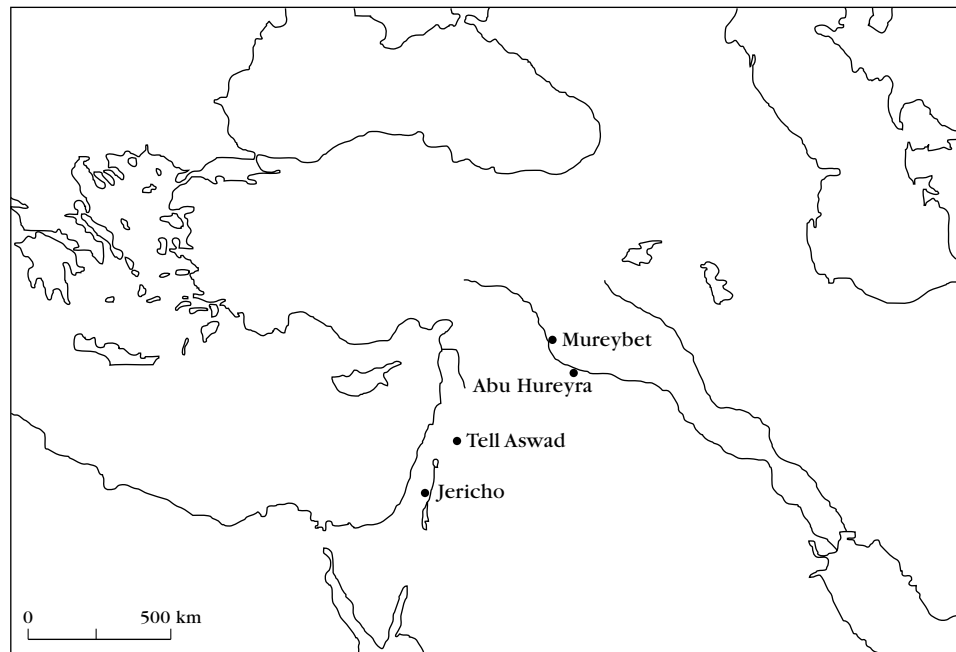
disappeared from memory many millennia in the past, although some farming peoples still recount tales of how they received other cultivated plants from gods, animate spirits, heroic ancestors, or the earth itself. But today we must use botanical and archaeological evidence to trace the story of wheat's domestication (implying a change in a plant's reproduction, making it dependent on humans) and its spread.

Domesticated wheats belong to at least three separate species (Zohary 1971: 238) and hundreds of distinct varieties, a number that continues to increase because the domestication of wheat continues. All domesticated wheat has lost the physical and genetic characteristics that would allow it aggressively to reseed and sprout by itself – losses which clearly distinguish domesticated wheats from wild relatives. Furthermore, both the remarkable geographic distribution of domesticated wheat and

the species' very survival depend on human beings. If no one collected wheat seeds and planted them in cleared, fertile ground, waving fields of grain soon would support hundreds of weeds, to be replaced by wild plants, and perhaps eventually by saplings and forests. Domesticated wheat and humans help each other in a relationship known as “mutualism” (Rindos 1984: 255).

Although humans domesticated wheat, one may argue that dependence on wheat also domesticated humans. The switch from gathering food to producing food, dubbed the “Neolithic Revolution” by V. Gordon Childe (1951: 74, orig. 1936), ultimately and fundamentally altered human development. Both wheat and barley, destined to feed the great civilizations of Mesopotamia, Egypt, Greece, and Rome, originated in the Near East, the earliest cradle of Western civilization (Map II.A.10.1). And with food production came great social and technological innovations. For example, because cereals can be stored year-round, early farmers could settle together in larger groups during the seasons when low food availability formerly had forced hunter-gatherers to disperse into small groups. Furthermore, by producing a surplus of cereal food, farmers could support others – people with specialized crafts, administrators, religious castes, and the like. Thousands of years later, cities emerged and empires arose. Clearly, the domestication of a cereal that has fed the Western world (and much of the rest of the world as well) holds a special place in the study of the origins of our foods.

Map II.A.10.1. The Ancient Near East showing sites mentioned in the text.



The Origins of Wheat and Barley Agriculture

While archaeologists recognize the momentous developments set in motion by food production in the ancient Near East, they continue to debate the essential factors that first caused people to begin farming wheat and barley.¹ How did agriculture begin? Which people first domesticated plants? Why did they do so when they did? And why did farming begin in only a few places? Answers to these questions are significant because with the domestication of wheat, humankind began the shift from hunting and gathering food to producing it. This change in lifestyle set humans on a new evolutionary course, and their society and environment were never the same after farming was established. Because wheat was one of the first crops to be farmed, its role in this fundamental shift has attracted much study, resulting in a variety of models to explain the process of domestication and its causes in the Near East.

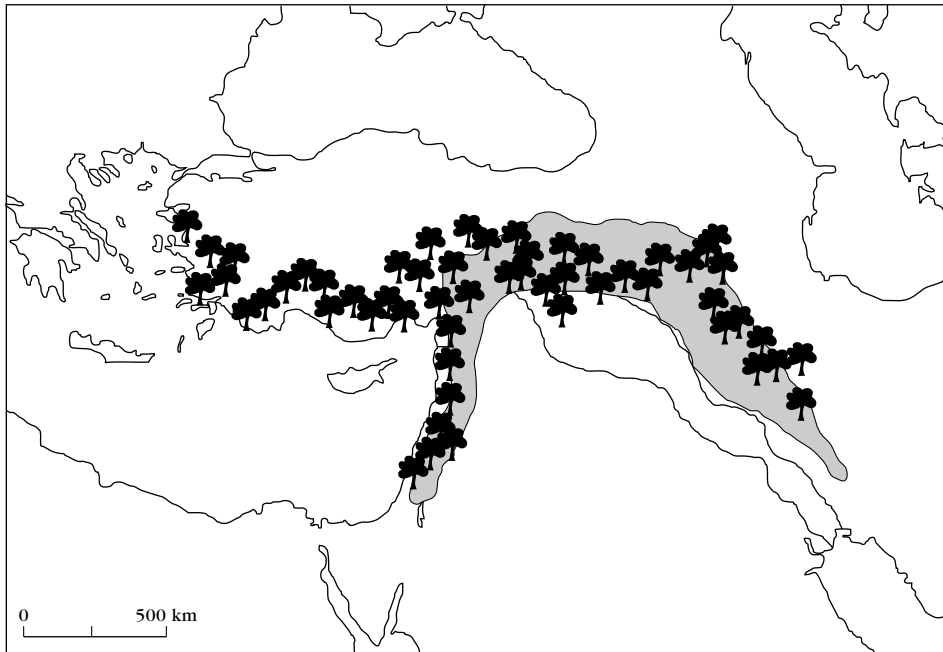
The Process of Cereal Domestication

To domesticate wheat, humans must have manipulated wild wheats, either through selective gathering or deliberate cultivation, with the latter implying activities such as preparing ground, sowing, and eliminating competing plants. We owe much of our understanding of the details of this process to the work of several botanists and archaeologists. A Russian botanist, Nikolai Vavilov (1951), for example, discovered that the greatest diversity in the gene pool of

wild wheats and barleys is in Southwest Asia. Where diversity is greatest, plants have been growing, fixing mutations, and interbreeding longest. Thus, it can be concluded that Southwest Asia was the ancestral homeland of these plants (Zohary 1970a: 33–5), and subsequent searches for the first farmers have concentrated on this region.

Robert Braidwood, an archaeologist from the University of Chicago, further refined the criteria for the homeland of wild wheats by identifying their modern ecological range, as the semiarid Mediterranean woodland belt known as the “hilly flanks” of the Fertile Crescent (Braidwood 1960: 134) (Map II.A.10.2). He reasoned that prefarming peoples had adapted culturally and ecologically to specific environments over long periods (a process he dubbed “settling in”) and that the first wheat farmers had already been living among natural stands of wild wheat.

One of Braidwood’s students and a great archaeologist in his own right, Kent V. Flannery, advocated explaining the origins of agriculture in terms of the process of plant domestication. His major contribution to modeling wheat domestication in the Near East stemmed from his recognition that plant domestication may have been the final result in a subtle chain of events that originated with changes in human food procurement occurring much earlier than the actual transition to agriculture (Flannery 1969, 1973: 284). Flannery’s “Broad Spectrum Revolution” portrays a shift in human ecology whereby humans began exploiting many previously minor food



Map II.A.10.2. The Near East with modern “hilly flanks” and Mediterranean woodlands.

sources such as cereals, fish, small game, and water fowl. Ultimately they came to depend on these food sources (1969: 79).

Flannery particularly emphasized the importance of moving cultigens (manipulated but not necessarily domesticated plants) on which humans depended “to niches to which [they were] not adapted” (Flannery 1965: 1251; Wright 1971: 460; Rindos 1984: 26–7). Thus, as people relocated to accommodate shifting population densities (Binford 1968: 332), they attempted to produce rich stands of cereals outside their natural range (Flannery 1969: 80). Such an effort would have helped domesticate wild wheats by preventing relatively rare genetic variants from breeding with the large pool of wild wheat types growing in their natural ranges.

David Rindos (1984) has described the general process of plant domestication from an evolutionary perspective, using the principles of natural selection and mutualistic relationships (coevolution) to describe how cereals, for example, would have lost their wild characteristics. In mutualistic relationships, domesticates and humans enhance each other's fitness, or ability to reproduce. In the case of wheat, women and men began to collect wild seeds in increasing quantities for food, while at the same time inadvertently selecting and replanting seeds from the plants best suited to easy harvesting (Harlan, de Wet, and Price 1973: 311; Kislev 1984: 63). Within a few generations, cultivated wheat plants became dependent on the harvesting process for survival, as wild self-planting mechanisms disappeared from the traits of cultivated wheats (Wilke et al. 1972: 205; Hillman and Davies 1990). The elimination of wild reseeding characteristics from a plant population ultimately accounted for the domestication of wheats.

The Causes of Cereal Domestication

Because the first evidence for agricultural societies occurs at the beginning of the Holocene (our present epoch) after a major climatic change, several archaeologists found a climatic explanation for the origins of agriculture to be plausible. Childe, for example, maintained that agriculture began at the end of the Pleistocene when climate change caused a lush landscape to dry up and become desert. Populations of humans, animals, and plants would have been forced to concentrate at the few remaining sources of water: Their enhanced propinquity would have provided increased opportunity for experimentation and manipulation (Childe 1952: 23, 25). Childe believed that agriculture started in the oases of North Africa and Mesopotamia, and although these locations were probably incorrectly targeted, some of his other hypotheses now seem essentially correct (Byrne 1987; McCorriston and Hole 1991: 60).

Lewis Binford (1968: 332–7) also indicated the importance of climate change when he emphasized the resource stress experienced by permanently set-

tled coastal populations hard-pressed by rising sea levels (also the result of climatic changes at the end of the Pleistocene). He pointed out, however, that population pressure in marginal zones (settled when rising sea levels flooded coastlines and forced populations to concentrate in smaller areas) would “favor the development of more effective means of food production” from lands no longer offering ample resources for scattered hunter-gatherers (1968: 332). Mark Cohen (1977: 23, 40–51) suggested that population growth filled all available land by the end of the Pleistocene; such dense populations eventually would have experienced the population pressure envisioned by Binford.

These ideas, however, only sharpened the question of why agriculture emerged in only a few regions. Accordingly, several archaeologists sought prerequisites – social or technological developments – that may have caused certain “preadapted” groups of hunter-gatherers to adopt farming in the Near East (Hole 1984: 55; Bar-Yosef and Belfer-Cohen 1989: 487; Rosenberg 1990: 409; McCorriston and Hole 1991: 47–9). Robert Braidwood thought that agriculture appeared when “culture was ready to achieve it” (Braidwood in Wright 1971: 457). For example, sedentism, which appeared for the first time just prior to agriculture (Henry 1985: 371–4, 1989: 219), would have profoundly affected the social relations in a group. Sedentary people can store and safeguard larger amounts of food and other goods than can mobile people. Stored goods increase the possibility of prestige being accorded to a relatively few individuals, since more opportunities now exist for redistribution of surplus goods through kinship alliances – the more goods a person distributes to dependents, the greater his prestige (Bender 1978: 213). As we have seen with contemporary sedentary hunter-gatherers, competition between leaders for greater alliance groups arguably stimulates an intensification of productive forces, which in turn provides a “major incentive for the production of surplus” (Godelier 1970: 120; Bender 1978: 213–14, 1981: 154). Perhaps wheat was such a desired surplus.

In parts of Southwest Asia, where sedentism appears to have preceded the development of agriculture (Henry 1981, 1989: 38–9; Bar-Yosef and Belfer-Cohen 1989: 473–4; Moore 1991: 291), it may have been the case that the causes of sedentism also contributed to the shift to food production (Moore 1985: 231; Henry 1989; Watson 1991: 14). On the other hand, Michael Rosenberg (1990: 410–11) argues that increasingly sharper territorial perceptions were the consequences of concentrated resource exploitation in such territories by hunter-gatherer groups already committed to mutualistic exploitation of plant resources.

A combination of factors probably best explains the domestication of cereals and the shift to agriculture in the Near East. Paleoenvironmental evidence

indicates that forests widely replaced a drier steppic cover (van Zeist and Bottema 1982). This has prompted Andrew Moore (1985: 232) to suggest that improved resources (resulting from climatic factors) enabled hunter-gatherers to settle; afterward their populations grew in size, so that ultimately they experienced the sort of resource stress that could have led to intensive manipulation of plants. Donald Henry (1985, 1989) also credits climatic change, several thousand years before agriculture emerged, with causing greater availability of wild cereals, which led increasingly to their exploitation. With this came dependence in the form of a sedentary lifestyle near wild cereal stands, and ultimately domestication during an arid spell when resources grew scarce.

Ecological factors play a major role in another combination model, in which the adaptation of wild cereals to a seasonally stressed environment is viewed as an explanation for the rise of agriculture in the Near East. Herbert Wright, Jr. (1977) was the first to recognize an association between the hot, dry summers and mild, wet winters that characterized Mediterranean climates and the expansion of wild, large-seeded, annual cereals. He suggested that these plants were absent from Southwest Asia in the late Pleistocene. Modern climatic models, in conjunction with archaeological evidence and ecological patterns, distinctly point to the southern Levant – modern Israel and Jordan – as the region where wheat farming first began 10,000 years ago (COHMAP 1988; McCorriston and Hole 1991: 49, 58). There, as summers became hotter and drier, plants already adapted to survive seasonal stress (summer drought), including the wild ancestors of wheat and barley, spread rapidly as the continental flora (adapted to cooler, wetter summers) retreated. Some hunter-gatherer groups living in such regions also experienced seasonal shortages of their erstwhile dependable plant resources. One group, the Natufians, named after the Wadi an Natuf in Israel (where archaeologists first found their remains) probably compensated for seasonal stress by increasingly exploiting the large-seeded annual wild wheats and barleys.

These various approaches, spanning nearly 50 years of research in the Near East, have all contributed to an increasingly sophisticated appreciation of the causes and the process of wheat domestication. Based on

data that either fit or fail to fit various models, many specific refutations appeared for each model, but these lie beyond the scope of this chapter. In addition, opinions on the process of domestication in the Near East still differ as follows:

1. Was the process fast or slow (Rindos 1984: 138–9; Hillman and Davies 1990: 213)?
2. Did domestication take place once or on many independent occasions (Ladizinsky 1989: 387; Zohary 1989: 369; Blumler 1992: 100–2)?
3. Was domestication primarily the result of biological processes (Binford 1968: 328–34; Flannery 1969: 75–6; Cohen 1977; Hayden 1981: 528–9; Rindos 1984) or the product of social changes (Bender 1978)?
4. Can the domestication process be linked to major changes in the global ecosystem (Childe 1952: 25; Binford 1968: 334; Wright 1977; Byrne 1987)?

Most archaeologists now believe that a complex convergence of multiple factors (climatic changes, plant availability, preadaptive technology, population pressure, and resource stress) accounts for the emergence of agriculture 10,000 years ago in the southern Levant (Hole 1984: 55; Moore 1985; Henry 1989: 40–55, 231–4; McCorriston and Hole 1991: 60; Bar-Yosef and Belfer-Cohen 1992: 39). However, there is still little consensus regarding the rapidity of the shift or the importance to be accorded to any single factor.

Archaeological Evidence for the Domestication of Wheat

The Evidence

The earliest remains of domesticated plants are the charred seeds and plant parts found on archaeological sites that date to the beginning of the Neolithic period. Unfortunately, other evidence for the use of plant foods in the past rarely shows exactly which plants were eaten. For example, grinding stones, sickles, and storage pits all indicate increased plant use and storage during the Early Neolithic period (about 10,000 to 8,000 years ago) (Table II.A.10.1), but they do not indicate which plants were processed (Wright 1994). In fact, such artifacts could have been used to process many kinds of plants and plant tissues, including many grasses, reeds, nuts, and tubers.

Table II.A.10.1. *Prehistoric cultures of the Near East*

Date	Period	Economy	Material culture
12,500–10,200 B.P.	Natufian	Hunting, gathering plants, and perhaps cultivating wild cereals	Grinding stones, storage pits, and sickles
10,200–9600 B.P.	Prepottery Neolithic A (PPNA)	Farming domesticates and hunting	Sickle blades, mudbrick architecture, axes, larger villages
9600–7500 B.P.	Prepottery Neolithic B (PPNB)	Farming domesticates and herding domesticated animals	Lime plaster, polished axes

Following the discovery of Neolithic crop plants (Hopf 1969), archaeologists have employed many analytical techniques to determine (1) whether even earlier peoples also cultivated plants, (2) whether such earlier uses of plant resources would have resulted in domestication, and (3) whether the first farmers originated in the region(s) where domesticated wheat first was found. The ultimate aim of such a quest, of course, has been to resolve the question of whether the earliest charred remains of domesticated wheat actually indicate the first wheat farmers.

In aiming at an answer, one must know the plant resources used by preagrarian hunter-gatherers and how the cultivation practices of the first farmers differed from plant use by their predecessors (Hillman 1989; Hillman, Colledge, and Harris 1989: 240–1). This knowledge, however, has proved elusive, largely because most direct evidence for prehistoric human use of plants has decayed or disappeared. Tools and pits may have been used for processing a wide range of plants. Chemical residues that allow archaeologists to specify *which* plants were eaten during the Early Neolithic period and the preceding Natufian period seldom have been preserved or examined (Hillman et al. 1993). Microscopic studies of the sheen left on flint sickle blades indicate that peoples using these tools reaped cereals (Unger-Hamilton 1989; Anderson 1991: 550), although it is impossible to ascertain which species.

Chemical composition of human bone also provides limited clues to plant consumption. For example, the ratio of strontium to calcium (Sr/Ca) in Natufian and Neolithic skeletons indicates that some early farmers eventually relied more heavily on animal foods than did their immediate Natufian predecessors (Sillen 1984; Smith, Bar-Yosef, and Sillen 1984: 126–8; Sillen and Lee-Thorp 1991: 406, 408). None of these isotopic data, however, have come from the very first farming populations (Pre-pottery Neolithic A); furthermore, such analyses cannot identify the specific plants that the first farmers ate.

The Sites

Neolithic sites with remains of domesticated wheat and other crops are the earliest known farming sites. But practices known to Neolithic farmers surely existed among their Natufian predecessors (Unger-Hamilton 1989) who for the first time in human history used large amounts of cereal processing equipment – grinding stones, sickle blades, storage pits – and lived year-round on one site (Bar-Yosef and Belfer-Cohen 1989: 468–70; Henry 1989: 195, 211–14, 219; Tchernov 1991: 322–9). Yet none of the Natufian sites excavated thus far have revealed domesticated wheat.

Furthermore, the presence of domesticated plants on Neolithic sites, more than any other evidence, has

defined our perception of a major economic difference between the first Neolithic farmers and their hunter-gatherer predecessors. Natufians may indeed have cultivated cereals, although they never apparently domesticated them, and traditions of cereal cultivation in conjunction with other gathering and hunting strategies probably persisted long into the Neolithic era when cereal farmers shared the Near East with other groups of people who were not especially committed to cultivation.

A few exceptional excavations have recovered plant remains from pre-Neolithic sites, but most of these have not yet been fully analyzed. The site of Abu Hureyra, along the banks of the Middle Euphrates River in northern Syria, yielded an abundance of charred plant remains reflecting the harvest of many types of wild seeds and fruits: These were gathered primarily in the local environments of the Late Pleistocene – steppe and steppe-forest, wadi banks, and the Euphrates River valley bottom (Hillman et al. 1989: 258–9).

The plant economy of Abu Hureyra's Epipaleolithic hunter-gatherers, however, does not appear to have led directly to farming. The site was abandoned at the time when farming began elsewhere (Moore 1975: 53, 1979: 68), and the evidence for a wide diversity of plants without evidence of intensive use of any particular one (Hillman et al. 1989: 265) is inconsistent with most models of cereal domestication (e.g., Harlan 1967; Rindos 1984; Henry 1989: 55, 216–17, 228; Hillman and Davies 1990: 212).

Instead, most models assume that cereal domestication followed intensive cereal exploitation by hunter-gatherers. At about the time that people abandoned Abu Hureyra, the sedentary inhabitants of nearby Tell Mureybet began to harvest two-seeded wild einkorn wheat and wild rye with unprecedented intensity (van Zeist and Bakker-Heeres 1984: 176–9; Hillman et al. 1993: 106). Although this type of wild wheat never developed into a domesticated plant (Zohary 1971: 239; van Zeist 1988: 58), the pattern of intensive cereal use at Tell Mureybet mirrors the type of economic pattern suggested for the Natufians from the southern Levant, where no plant remains from Epipaleolithic sites have been fully analyzed (Colledge 1991).

The southern Levant is where the earliest domesticated wheat appears. In the period known as the Pre-pottery Neolithic A (approximately 9,000 to 10,000 years ago²), the early farming site of Jericho (in the Jordan Valley) has two types of domesticated wheat grains, einkorn and emmer (Hopf 1969: 356, 1983: 581) (Table II.A.10.2). Some of the oldest dates from Jericho can be questioned (Burleigh 1983: 760; Bar-Yosef 1989: 58), and domesticated wheat seeds from Jericho may actually be several hundred years younger than the oldest Neolithic radiocarbon dates (10,500–10,300 years ago) suggest.

Table II.A.10.2. *Principal wheat types*

Wheat (<i>Triticum</i>) types					
Botanical name	English name	Ploidy	Rachis	Glumes	Remarks
<i>T. boeoticum</i> var. <i>aegilopoides</i>	Wild einkorn	2x ^a AA	Brittle	Tight ^b	Spikelets 1-grained, ancestor of einkorn wheat, modern range in Taurus mts. ^c
<i>T. boeoticum</i> var. <i>tbaoudar</i>	Wild einkorn	2x AA	Brittle	Tight	Spikelets 2-grained, collected wild in northern Levantine Neolithic, ^d modern range in western Anatolia ^e
<i>T. monococcum</i>	Einkorn	2x AA	Tough	Tight	Domesticated primitive wheat
<i>T. dicoccoides</i>	Wild emmer	4x AABB	Brittle	Tight	Ancestor to emmer, modern range is basalt uplands of Syria, Jordan, Israel, and Taurus ^f
<i>T. dicoccum</i>	Emmer	4x AABB	Tough	Tight	Most favored wheat of the ancient world, widely cultivated, India-Britain
<i>T. durum</i>	Macaroni wheat	4x AABB	Tough	Free	Widely used for pasta, derived from emmer
<i>T. turgidum</i>	Rivet/cone	4x AABB	Tough	Free	Recent (16th C.) species (like <i>T. polonicum</i> , 17th C.), derived from macaroni wheat, occasionally branched spikelets
Many other varieties/species		4x AABB	Tough	Free	
<i>T. timopheevii</i>	Timopheevii wheats	4x AAGG	Tough	Free	Group of allotetraploids sharing only 1 genome with emmer and durum wheats; they arose independently in eastern Turkey ^g
<i>T. aestivum</i>	Bread wheat	6x AABBDD	Tough	Free	Major modern cereal crop widely grown; gluten, a sugar, allows yeast to reproduce, thus dough made from this flour rises; must appear after tetraploid wheats (see also <i>T. spelta</i>)
<i>T. spelta</i>	Spelt	6x AABBDD	Brittle	Tight	Range in northern Europe, possibly preceded bread wheat, ^h only a relic crop today
No wild hexaploid wheats					
<i>T. speltooides</i>	Goat-faced grasses	2x BB	Brittle	Tight	Probably contributed half the chromosomes of wild emmer
<i>T. tauschii</i> (= <i>Aegilops squarrosa</i>)	Goat-faced grasses	2x DD	Brittle	Tight	Contributed gluten and cold-hardiness to crosses with tetraploid wheats, modern distribution Central Asia and Transcaucasia

^aPloidy refers to the number of chromosome sets. Diploid plants have 2 sets of chromosomes, whereas tetraploids (4 sets) may arise, as in the case of some wheats, when different diploid plants cross to produce fertile offspring that carry chromosome sets from both parents. Hexaploids (with 6 sets) arise from the crossing of a diploid and tetraploid.

^bGlumes adhere tightly to grain, protecting it from predators and spoilage. This wild characteristic is not lost until the appearance of free-threshing wheats (with loose glumes easily releasing grains), such as macaroni and bread wheat. Loose glumes apparently were secondarily selected for, since the wheats with this characteristic ultimately derive from glume wheats, such as emmer (Zohary 1971: 240, 243).

^cHarland and Zohary 1966; Zohary 1971: 239; van Zeist 1988: 54.

^dvan Zeist 1970: 167–72; van Zeist and Bakker-Heeres 1984 (1986): 183–6, 198; Watkins, Baird and Betts 1989: 21.

^eHarlan and Zohary 1966; Zohary 1971: 239; van Zeist 1988: 54.

^fHarlan and Zohary 1966; Zohary 1971: 240; Limbrey 1990. One significant advantage of allotetraploidy in plants (chromosome pairs inherited from two ancestor plants) is that the additional genome often increases ecological tolerance in plants so that allotetraploids may occupy the geographic ranges of both parent plants as well as their overlap (Grant 1981).

^gLilienfeld 1951: 106, but see Zohary 1971: 241–2.

^hZeven (1980: 31) suggests that the expected evolutionary path of wheats would have emmer (a glume wheat) cross with a wild grass (also with tight glumes) to produce spelt wheat. The same wild grass would later cross with durum derived from a mutant emmer strain to produce bread wheat. Mitigating against this scenario is the very early appearance of bread wheat in archaeological sites (Helbaek 1966) and the genetic evidence suggesting that free-threshing characters easily and quickly may become fixed in a population (Zohary and Hopf 1988: 46). Bread wheat probably quickly followed the appearance of hexaploid spelt wheat.

Sources: Kimber and Feldman (1987), Lilienfeld (1951), Percival (1921), Zeven (1980), Zohary (1971), Zohary and Hopf (1988).

Today Jericho lies at the edge of a spring whose outflow creates an oasis in the arid summer landscape of the Jordan Valley. This alluvial fan, created by winter streams flowing from the Judean hills, nourishes palms and summer crops in the midst of a shrubby wasteland, but the area looked different during the Early Neolithic. Most of the sediment accumulated around the site of the early farming village at Jericho has washed downslope since the Neolithic (Bar-Yosef 1986: 161), perhaps because the shady glades of wild trees – pistachio, fig, almond, olive, and pear (Western 1971: 36, 38; 1983) – were stripped from the surrounding hillsides thousands of years ago. The Neolithic inhabitants planted some of the earliest wheat ever farmed, and they depended on the supplemental water provided by the spring and flowing winter streams to ensure their harvests.

The farmers at Jericho necessarily managed water: Floods frequently threatened to damage their habitations and storage areas. They built terrace walls and dug ditches to divert the flow (Bar-Yosef 1986: 161; compare Kenyon 1979: 26–7) from their small round and lozenge-shaped houses with their cobble bases and mud-brick walls (Kenyon 1981: 220–1). Their apparent choice of supplementally watered land to grow wheat was unprecedented, for wild wheats had hitherto thrived on dry slopes at the edge of Mediterranean forest (Limbrej 1990: 46, 48).

The only other site that has yielded domesticated wheat from the same general era as Jericho is the site of Tell Aswad about 25 kilometers southeast of Damascus, Syria (Contenson et al. 1979; Contenson 1985). In the earliest midden layers of this prehistoric settlement, along the margins of a now-dried lake, archaeologists recovered domesticated emmer wheat along with barley and domesticated legumes such as lentils and peas. Any former dwellings had long since been destroyed, perhaps because structures consisted largely of wattle (from reeds) and daub (Contenson et al. 1979: 153–5).

Today Tell Aswad lies outside the green Damascus oasis on a dusty, treeless plain occupied by the modern international airport, but its former setting was quite different. We know from charred seeds of marshy plants, historical accounts of the environment (van Zeist in Contenson et al. 1979: 167–8), and pollen studies (Leroi-Gourhan in Contenson et al. 1979: 170) that the lake once adjacent to the site was much larger; in addition, there were many wild trees adapted to a semiarid Mediterranean forest-steppe (pistachios, figs, and almonds). Pollen of species such as myrtle and buckthorn (*Rhamnus* spp.) may indicate rainfall greater than the annual 200 millimeters today (Leroi-Gourhan in Contenson et al. 1979: 170). Under wetter conditions, farmers were probably able to grow wheat and other crops. When it was drier, they probably used the extra moisture afforded by the lake and autumn flooding to grow wheats beside the lake shores.

Tell Aswad and Jericho are critical sites in the history of wheat agriculture. To be sure, we cannot be

certain that the farmers who settled at the edge of Lake Ateibe (Tell Aswad) and near the spring feeding into the Jordan Valley at Jericho were the first people ever to grow domesticated wheat because archaeologists will never know for certain if earlier evidence awaits discovery elsewhere.

It is interesting to note, however, that contemporary evidence in adjacent regions suggests that people had not domesticated plants by 8000 B.C. In the Nile Valley of Egypt, for example, farming appears much later, around 5000 B.C. (Wenke 1989: 136), and in northern Syria on such early settlements as Tell Mureybet, people exploited wild, not domesticated, wheats and rye (van Zeist 1970: 167–72; van Zeist and Bakker-Heeres 1984: 183–6, 198; Hillman et al. 1993: 106). Recent research in the Taurus Mountains of southeastern Turkey has focused on early settled communities that apparently were not intensively exploiting wild or domesticated cereals (Rosenberg et al. 1995). Southern Mesopotamia, where the first cities emerged, saw agricultural settlements only in later times (Adams 1981: 54), and the surrounding mountains continued to support pastoralists and hunter-gatherers long after farming appeared in the southern Levant.

Botanical Evidence

Taxonomy

Botanical and ecological evidence for the domestication of wheat and its differentiation into many species also partially contributes to an understanding of where and when the first domestication occurred. Many different morphological forms of wheat appear in the archaeological record, even as early as the Neolithic deposits at Jericho, Tell Aswad, and Tell Mureybet along the northern Euphrates River (van Zeist 1970: 167–72; van Zeist and Bakker-Heeres 1984: 183–6, 198). The different forms of wild and domesticated wheats are of incalculable value to archaeologist and botanist alike, for these different plants that can be distinguished from archaeological contexts allow botanists and ecologists to identify wild and domesticated species and the conditions under which they must have grown.

The forms recognized archaeologically, moreover, may not always conform to wheat classification schemes used by modern breeders and geneticists. Wheat classification is complex and confusing, for hundreds of varieties have appeared as wheat farming spread around the world. Although many different kinds of wheat can be readily distinguished by their morphological characteristics (such as red or black awns, hairy glume keels, spikelet density), other varieties can cross-fertilize to combine characters in a perplexing array of new plants. The great variability in the visible characteristics of wheats has led to confusion over how to classify different species – a term employed in its strictest sense to describe reproductively isolated organisms (Mayr 1942; Baker 1970: 50–1,

65–6). In the case of wheat, many botanists commonly accept as distinct species morphologically distinct types that can readily cross to form fertile hybrids with other so-called species (Zohary 1971: 238).

Because botanists rely on both morphological and genetic characteristics to identify different wheats, classificatory schemes (of which many exist, for example, Percival 1921; Schiemann 1948; Morris and Sears 1967; Löve 1982) must take both aspects into account (Zohary 1971: 236–7; but compare Baker 1970). Using morphological traits, taxonomists originally split wild and cultivated wheats into at least a dozen different taxa, many of which are highly interfertile. Geneticists, however, maintain that all domesticated wheats belong to four major groups that produce only sterile crosses; furthermore, they include the wild grass genus, *Aegilops*, in the *Triticum* genus, since several taxa of wild *Aegilops* contributed chromosome sets (genomes) to domesticated wheats by crossing with wild wheat plants (Zohary 1971: 236).

Many of the wheats distinguished by taxonomists, however, lose their identifying genetic signatures when charred, abraded, and preserved for thousands of years in archaeological sites. Because fragile genetic material only recently has been demonstrated to have survived this process (Brown et al. 1993), morphological features that can be used to distinguish different wheat

types have made traditional taxonomic schemes (based on morphology) of great value to archaeologists. Furthermore, some of the major behavioral characteristics of cultivated and wild wheats do have morphological correlates that endure in the archaeological record. These features also reflect significant events in the domestication of wheat (Figure II.A.10.1).

The most significant of these morphological features is rachis (segmented stem) durability. Wild wheats and wild *Aegilops*, a morphologically distinct grass genus with species capable of crossing with many wheats, have a rachis capable of shattering, once the grains have matured, into pieces bearing one or two grains. These pieces, or spikelets, taper at their bases and carry stiff hairs that act as barbs to facilitate the spikelets' entry into cracks in the soil. In wild wheats, grains are tightly enclosed in tough glumes that protect them from predation.

In domesticated wheats, these features vanish. The rachis fails to shatter when ripe, a feature particularly important to humans who harvest using sickles – the tools introduced by Natufian and early Neolithic groups (Hillman and Davies 1990: 172–7) (Figure II.A.10.2). In the relatively pure stands of wild wheats, at the margins of Mediterranean oak forests where agriculture began, harvesting methods would fundamentally affect the domestication process (Harlan

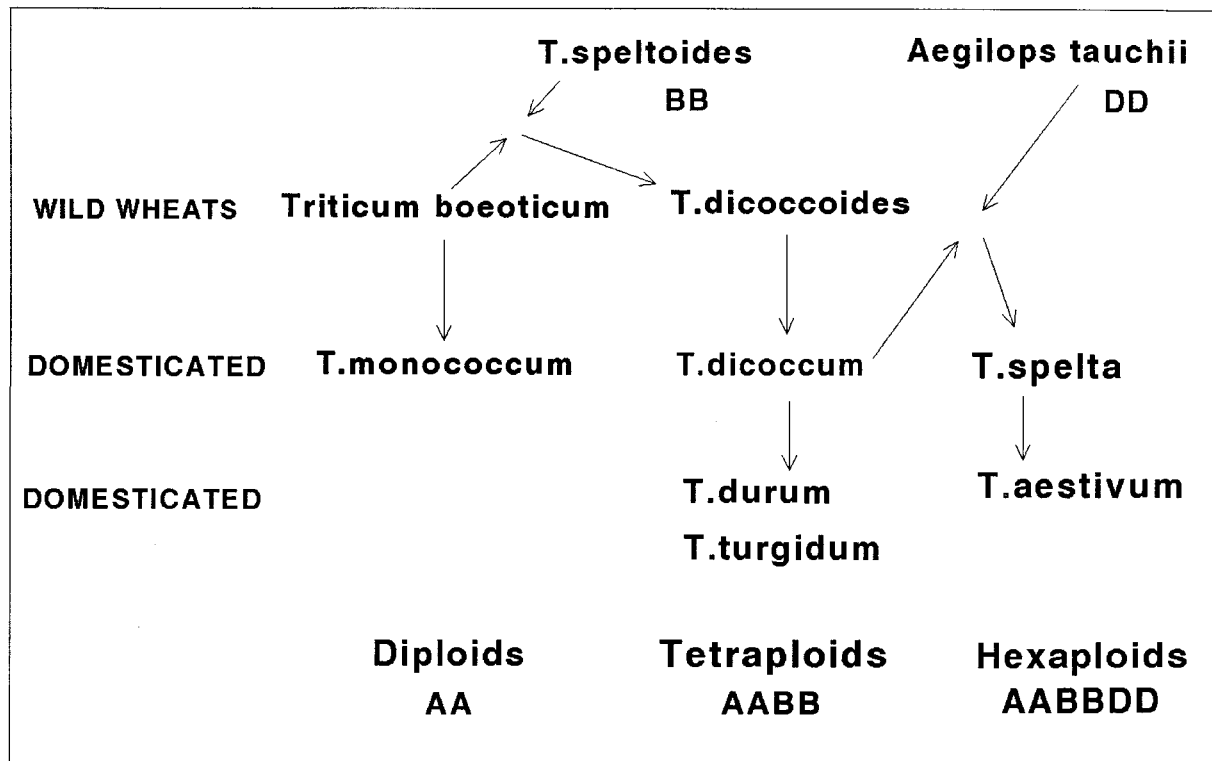


Figure II.A.10.1. Related wheats and goat-faced grasses. (After Zohary 1970b: 241; Hillman, personal communication, 1984).

1967, 1989; Bohrer 1972; Wilke et al. 1972: 205; Hillman and Davies 1990: 172-7). Harvesters use fairly violent motions when equipped with sickles or when uprooting plants to harvest straw and seed. These methods tend to shatter ripe ears, leaving for collection either immature seed (unfit for germination the following year if replanted) or relatively rare genetic mutants with tough rachises. Although these rare plants reproduce poorly in the wild, they are ideal for cultivation, as ripe seed can regenerate if replanted (Helbaek in Braidwood and Howe 1960: 112-13). By unconscious selection (Rindos 1984: 86-9) for a tough rachis gene, harvesters may replace a wild population with a domesticated one in as few as 20 to 30 years (Hillman and Davies 1990: 189).

Wild and domesticated cereals often can be distinguished when examining rachis fragments in archaeological plant remains (for example, Bar-Yosef and Kislev 1986; Kislev, Bar Yosef, and Gopher 1986: 198-9; compare Bar-Yosef and Belfer-Cohen 1992: 37-8). The earliest known domesticated wheats from Tell Aswad exhibit tough rachises (van Zeist and Bakker-Heeres 1982: 192-6). At the same period, the wheats intensively harvested along the Middle Euphrates River at

Tell Mureybet (van Zeist 1970: 167-72; van Zeist and Bakker-Heeres 1984: 183-6, 198) and in northern Mesopotamia at the site of Qeremez Dere (Watkins, Baird, and Betts 1989: 21) remained wild, perhaps partly because of a harvesting technique that favored the proliferation of brittle-rachis types in the population. For example, beating wild grass heads over baskets to collect seed was a technique widely employed in many parts of the world where no domestication occurred (Bohrer 1972: 145-7; Wilke et al. 1972: 205-6; Harlan 1989; Nabhan 1989: 112-18). Although baskets and wooden beaters have a low probability of surviving in archaeological sites in the Near East, the remarkable paucity of sickle blades at Tell Mureybet (Cauvin 1974: 59) would support a suggestion that people may have harvested wild cereals by a different method from that used at Jericho and Tell Aswad, where sickle blades are more common.

Cytogenetic Evidence

The results of modern genetic studies have also contributed incomparably to disentangling the history of domesticated wheat. In an effort to improve modern strains of bread wheat and to discover new genetic combinations, biologists have compared the genetic signatures of different varieties, types, and species of wheats. Genetic differences and similarities have allowed specialists to trace relationships among various forms of wild and domesticated wheats and to determine which wild wheats were ancestral to domesticates.

All of the relationships described in Figure II.A.10.1 and Table II.A.10.2 have been confirmed by genetic tests (Zohary 1989: 359). Of particular importance to domestication, the work of H. Kihara has largely defined the cytogenetic relationships between emmer, durum, and hexaploid wheats (Lilienfeld 1951). Domesticated emmer wheat shares close genetic affinities with its wild progenitor (*Triticum dicoccoides* = *Triticum turgidum* subsp. *dicoccoides*) and is largely a product of unconscious human selection for a tough rachis. Durum wheats and rivet wheats likewise received their 2 chromosome sets from wild emmer (Zohary 1971: 239) and probably are secondarily derived from domesticated emmer through selection for free-threshing characteristics, larger seeds, and various ecological tolerances (for example, Percival 1921: 207, 230-1, 241-3).

Hexaploid wheats, which belong in a single cytogenetic taxon (Zohary 1971: 238; Zohary and Hopf 1993: 24), have no wild hexaploid ancestors: They emerged as a result of a cross between domesticated tetraploid wheats (which may or may not have been free-threshing) and a wild grass native to continental and temperate climates of central Asia (Zohary and Hopf 1988: 46). This implies that hexaploid wheats emerged only when tetraploid wheats spread from the Mediterranean environment to which they were adapted. From archaeological evidence of the spread



Figure II.A.10.2. Photograph of the Nahal Hemar sickle. (Photo: M. Barazani-Nir, Centre de Recherches Français de Jérusalem, O. Bar-Yosef, and D. Alon.)

of farming, one assumes that hexaploid wheats appeared after 7500 B.C.³ True bread wheats with loose glumes probably came from spelt ancestors, but only two slight genetic changes produce loose glumes (Zohary and Hopf 1988: 46), implying that the mutations may occur easily and become rapidly fixed in a domesticated population.

Cytogenetic studies also have suggested that domestication occurred in only one population of wild wheats, from which all modern conspecific cultigens (of the same species) are derived. All the varieties and species of tetraploid wheats have the same basic genetic constitution as wild emmer wheat (AABB genomes)⁴ rather than timopheevii wheat (AAGG). This indicates that if multiple domestications had occurred, timopheevii wheat, which is morphologically indistinguishable from wild emmer, would have been systematically ignored. A more parsimonious explanation is that of Daniel Zohary (1989: 369), who suggests that emmer wheat was domesticated once and passed from farming community to community (see also Runnels and van Andel 1988). Archaeological evidence on Crete and in Greece (Barker 1985: 63-5) indicates that fully domesticated wheats were introduced to Europe from the Near East (Kislev 1984: 63-5). An alternative hypothesis - that hunter-gatherers in Europe independently domesticated emmer and einkorn from native wild grasses (Dennell 1983: 163) - has little supporting evidence. Botanists using cytogenetic evidence, however, may more easily recognize evidence for single domestication than for multiple events, genetic traces of which can be obscured by other biological and historical processes (Blumler 1992: 99, 105).

Ecology of Wheats

Perhaps it will never be possible to determine unequivocally whether wheat species were domesticated in one place or in several locations. Nevertheless, the ecological constraints limiting the growth of different species, varieties, and forms of wild and domesticated wheats narrow greatly the possibilities of where and under what ecological circumstances wheat may have been domesticated. Ecological constraints have been examined both on a macro and micro scale, and both scales contribute significantly to our understanding of wheat domestication.

On a macro scale, the geographic distributions of discrete species or varieties of wild wheats provide ecological ranges within which, or indeed adjacent to which, researchers locate wheat domestication and the origins of agriculture (Harlan and Zohary 1966) (Maps II.A.10.3-5). Using modern wild wheat distributions, botanists and archaeologists have singled out the southern Levant and Taurus range as the most likely source of domesticated emmer and einkorn (Harlan and Zohary 1966; Zohary 1971: 239-42), although bread wheats may have quickly evolved

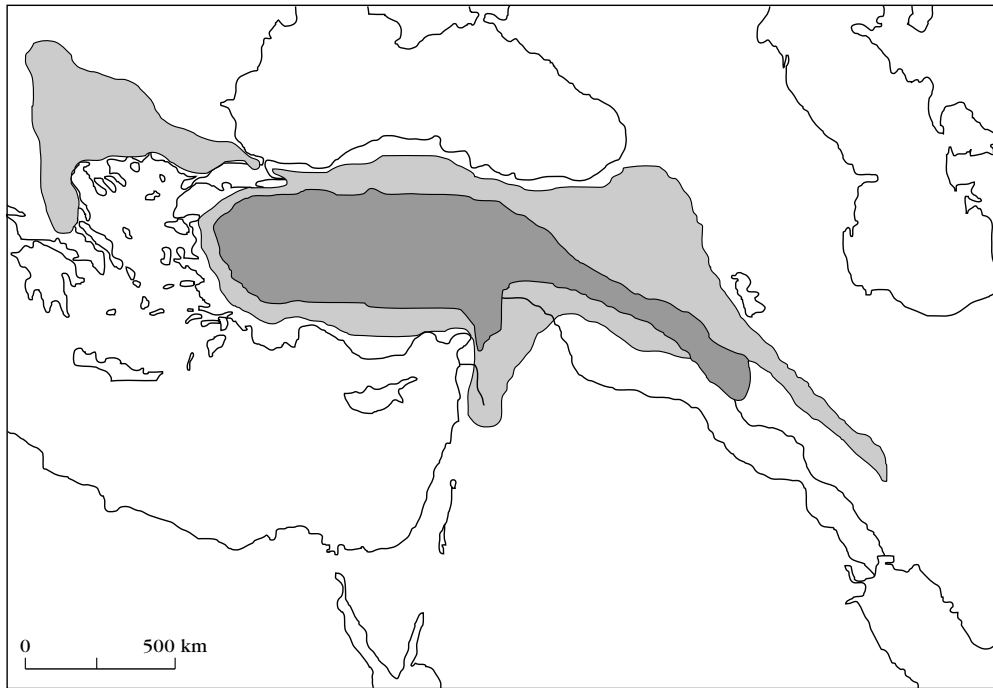
from spelt wheat somewhere in the Caspian region (Zohary 1971: 244; Zeven 1980: 32). Timopheevii wheats represent merely a later independent domestication of tetraploids in eastern Anatolia and Georgia. The conclusions of Vavilov, Braidwood, Flannery, Harlan, and D. Zohary depend greatly on modern geographical distributions of wild wheats.

Nevertheless, a serious problem in using the modern ranges of wild wheats is the assumption that these ranges reflect the former natural extent of wild species. In the past 10,000 years in the Near East, climates as well as human land use patterns have fluctuated. Grazing, deforestation, and suppression of natural forest fires have had a profound effect on vegetation (Naveh and Dan 1973; Naveh 1974; Le Houérou 1981; Zohary 1983), altering not only plant distributions but the character of entire vegetation zones (McCorrison 1992).

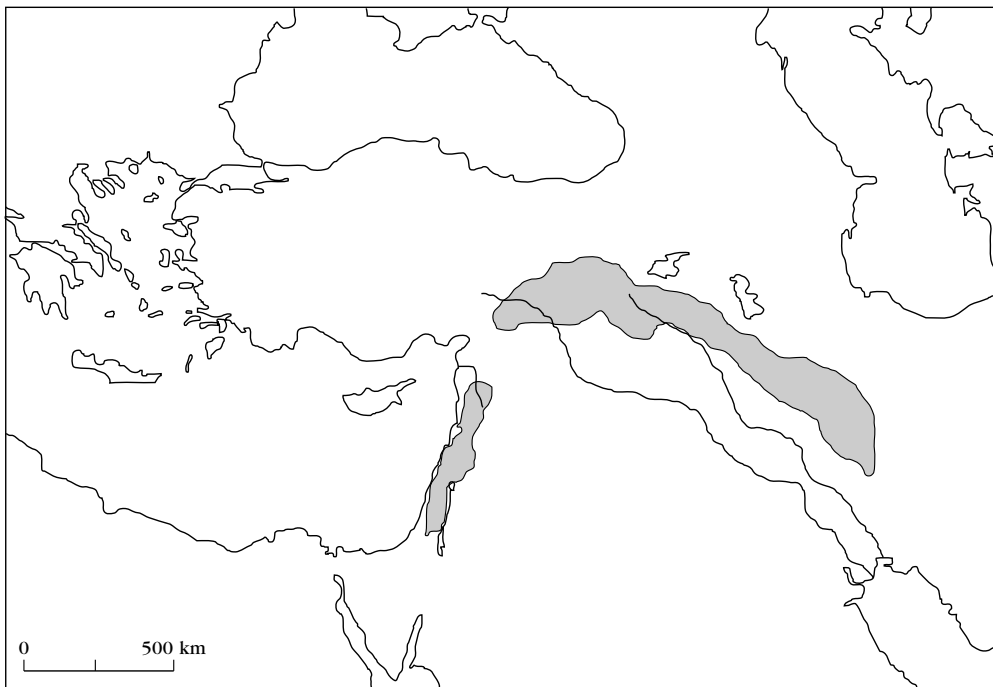
Water, light, and soil properties determine growth on a local scale within the geographic ranges of wheats. Wheat plants, like many other crops, require land free of competition from established plants that tap water and block light and where the seed may embed itself in the earth before ants or other predators discover it (Hillman and Davies 1990: 164; Limbrey 1990: 46).

Truly wild einkorn and emmer typically thrive on the open slopes at the margins of scrub-oak forests: They are poor competitors in nitrogen-rich soils typical of weedy habitats, such as field margins, habitation sites, and animal pens (Hillman and Davies 1990: 159, 160; Hillman 1991; McCorrison 1992: 217; compare Blumler and Byrne 1991). These latter sites readily support domesticated glume wheats (emmer and einkorn). Their wild relatives prefer clay soils forming on "basalt or other base-rich fine-grained rocks and sediments under warm climates with a marked dry season" (Limbrey 1990: 46). Indeed, some evidence suggests that wild wheats were first harvested from such soils (Unger-Hamilton 1989: 100; Limbrey 1990: 46).

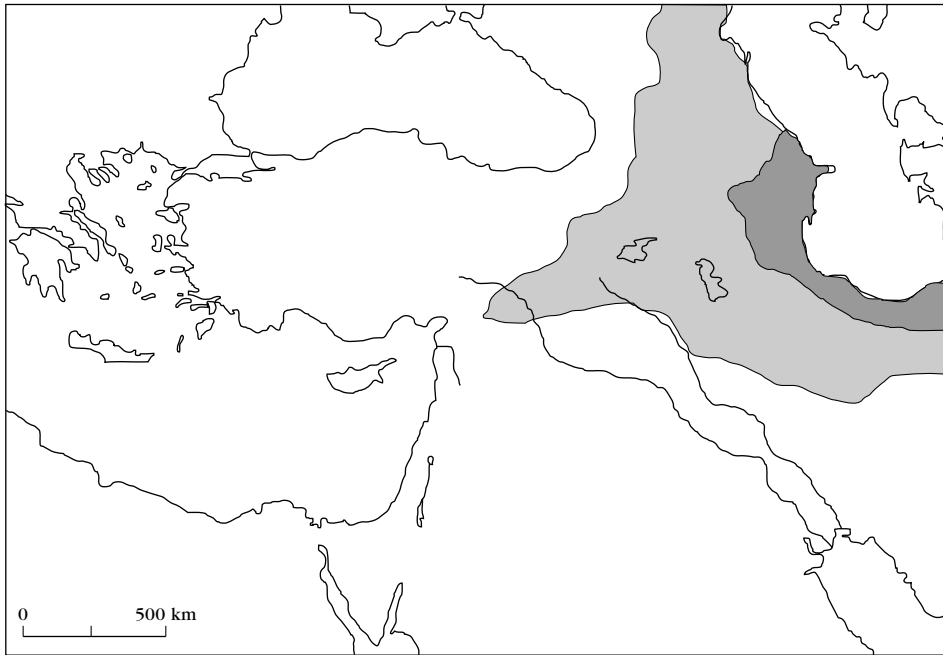
Early farming sites like Jericho, Tell Aswad, and Mureybet, however, were located beside alluvial soils in regions of low rainfall where supplemental watering from seasonal flooding and high water tables would have greatly enhanced the probability that a wheat crop would survive in any given year. If the wild wheats originally were confined to the upland basaltic soils, they must have been deliberately moved to alluvial fields in the first stages of domestication (Sherratt 1980: 314-15; McCorrison 1992: 213-24; compare Hillman and Davies 1990). Removal of a plant from its primary habitat often causes a genetic bottleneck (Lewis 1962; Grant 1981) whereby the newly established population is fairly homogeneous because its genetic ancestry comes from only a few plants. This tends greatly to facilitate domestication (Ladizinsky 1985: 196-7; Hillman and Davies 1990: 177-81).



Map II.A.10.3. Geographic distribution of wild einkorn wheat, *Triticum boeoticum*. (After Harlan and Zohary 1966; D. Zohary 1989.) Darker shading indicates nonweedy populations.



Map II.A.10.4. Geographic distribution of wild emmer wheat, *Triticum dicoccoides*. (After Harlan and Zohary 1966.)



Map II.A.10.5. Geographic distribution of goat-faced grass, *Aegilops tauchii*. (After Harlan and Zohary 1966.)

The Spread of Domesticated Wheats from the Near East

Deliberate planting of wheat in new habitats (supplementally watered alluvial soils) is one of only a few known events in the earliest spread of wheat farming. Archaeologists understand very poorly the processes and constraints that led to the spread of wheat into different environments and regions of the Near East. They understand far better the spread of agriculture in Europe because of research priorities set by Childe and other Western archaeologists who identified the arrival of domesticated wheat in Europe around 6000 B.C. (Barker 1985: 64; Zohary and Hopf 1988: 191) as an event of critical significance in the history of humankind (viewed from a European perspective).

Thus, in contrast with the Near East, in Europe the process of transition from different styles of hunting and gathering to a predominantly agricultural economy and the adaptation of crops such as wheat to new environments has received intense archaeological and theoretical consideration. The progress of Neolithic settlement across the plains of Greece and Italy, up the Balkans and into the river basins of central and eastern Europe, across the forested lands, and into the low countries encompasses many regional archaeologies and localized theoretical explanations. Wheat accompanied the Neolithic nearly everywhere in Europe, although by the time farming took hold in Britain around 3500 B.C., the cultivated varieties of einkorn, emmer, and spelt must have tolerated colder and longer winters, longer daylight during the ripen-

ing season, and greatly different seasonal rainfall than that of the Mediterranean lands from which the crops originated.

Wheat also spread out of the Near East to Africa, where it could be found in northern Egypt after 5000 B.C. (Wenke 1989: 136; Wetterstrom 1993: 203–13). With other introduced crops, wheat fueled the emergence of cultural complexity and replaced any indigenous attempts at agriculture initiated in subtropical arid lands to the south (Close and Wendorf 1992: 69). Because most wheats require cool winters with plentiful rain, the plant never spread widely in tropical climates where excessive moisture during growing and ripening seasons inhibits growth and spurs disease (Lamb 1967: 199; Purseglove 1985: 293). But the grain also spread to South Asia where as early as 4000 B.C. hexaploid wheats were cultivated at the Neolithic site of Mehrgarh (Pakistan) (Costantini 1981). By the third millennium B.C., the Indus Valley civilization, city-states in Mesopotamia, and dynastic Egypt all depended on domesticated wheat and other cereals.

In the sixteenth century, colonists from the Old World brought wheat to the New World: The Spanish introduced it to Argentina, Chile, and California where the cereal flourished in climates and soils that closely resembled the lands where it already had been grown for thousands of years (Crosby 1986; Aschmann 1991: 33–5). The political and social dominance of European imperialists in these new lands and the long history of wheat farming in the Old World – where crop and weeds had adapted to a

wide range of temperature, light, and rainfall conditions (Crosby 1986) – largely accounts for the fact that wheat is one of the world's most significant crops today. Its domestication is a continuing process, with yearly genetic improvements of different strains through breeding and new gene-splicing techniques (Heyne and Smith 1967).

Summary

Botanical and archaeological evidence for wheat domestication constitutes one of the most comprehensive case studies in the origins of agriculture. Some of the issues discussed in this chapter remain unresolved. For many of the domesticated wheats, including the primitive first wheats (einkorn and emmer), the integration of botanical and archaeological evidence indicates where and approximately when people entered into the mutualistic relationship that domesticated both humans and their wheats. Less is understood about the origins of hexaploid wheats, largely because the results of archaeological investigations in central Asia have long been inaccessible to Western prehistorians.⁵ From botanical studies, however, we can (1) trace the wild ancestors of modern wheats, (2) suggest that isolated events led to the domestication of each species, and (3) reconstruct the environmental constraints within which the first farmers planted and reaped.

Yet, some debate still continues over the question of how wheat was domesticated. Did people in many communities independently select easily harvested plants from weedy wild emmer growing on their dump heaps (Blumler and Byrne 1991)? Or did they, as most believe, deliberately harvest and sow fully wild wheats and in the process domesticate them (Unger-Hamilton 1989; Hillman and Davies 1990; Anderson-Gerfand, Deraprahamian, and Willcox 1991: 217)? Related questions, of course, follow: Did this latter process happen once or often, and did the first true farmers move these wild plants to seasonally flooded or supplementally watered soils as part of the domestication process?

Scholars also continue to discuss why people began to domesticate wheat in the first place some 10,000 years ago as part of the larger question of the origins of agriculture. Although most agree that many factors were involved in the Near East and that they converged at the end of the Pleistocene, there is no agreement about which factor or combination of factors was most important (for example, Graber 1992; Hole and McCorrison 1992). But because there are no written Neolithic versions of the original crop plants (compare Hodder 1990: 20–1), it is up to us – at the interdisciplinary junction of archaeology and botany – to continue to reconstruct this evolutionary and cultural process in human history.

Joy McCorrison

Endnotes

1. This review is summarily brief; for more thorough treatment, the reader should consult other sources, especially G. Wright 1971, Henry 1989, and Watson 1991. As theoretical explanations for domestication of wheats and barley overlap considerably, both are discussed in this chapter.
2. Neolithic dates are generally quoted as uncorrected radiocarbon dates because exact calendar dates are unknown. The basis for radiocarbon dates is a ratio of carbon isotopes in the earth's atmosphere, but this ratio varies at different times. For later periods, analysts correct for the variation in atmospheric carbon isotope ratios by comparing radiocarbon dates with exact dendrochronological (tree-ring) calendrical dates. Unfortunately, the sequence of old timbers necessary for tree-ring dates does not extend back as far as the Neolithic in the Near East, leaving archaeologists without a dendrochronological calibration scale to correct their radiocarbon dates. Although some progress has been made with isotopically dated coral reefs, there are still problems with variation in atmospheric carbon isotope ratios at different periods. Archaeologists should continue to quote uncalibrated dates.
3. This inference is conservatively based on the dates of proven Neolithic (PPNB) wheat farmers in the Taurus mountains and northeastern Syria and Iraq. Although both areas are poorly known archaeologically in the early Neolithic (Pre-pottery period, they fall within the modern range of the wild goat-faced grass (*Aegilops tauschii*). Since *A. tauschii* contributed genetic tolerance for cold winters (continental and temperate climates) the hexaploid wheats (Zohary, Harlan, and Vardi 1969; Zohary and Hopf 1993: 50), and since hexaploid wheats emerged after the domestication of tetraploids (Zohary and Hopf 1988: 46), the appearance of domesticated emmer in the PPNB (after approximately 7500 B.C.) (van Zeist 1970:10) serves as a *terminus post quem* for the domestication of hexaploid wheats.
4. Kihara defined as a genome "a chromosome set . . . a fundamental genetical [sic] and physiological system whose completeness as to the basic gene content is indispensable for the normal development of gonads [the special reproductive cells] in haplo- and zygotes in diplophase" (Lilienfeld 1951: 102). In diploid plants, the genome represents all chromosomes of the plant; in allotetraploids, the genome is derived from the chromosomes of both contributing ancestral plants.
5. A recently initiated research project seeks to clarify agricultural origins in the Caspian region at the site of Jeitun in the Kara Kum desert of Turkmenia (see Harris et al. 1993).

Bibliography

- Adams, Robert McC. 1981. *Heartland of cities*. Chicago.
- Anderson, Patricia C. 1991. Harvesting of wild cereals during the Natufian as seen from experimental cultivation and harvest of wild einkorn wheat and microwear analysis of stone tools. In *The Natufian culture in the Levant*, ed. Ofer Bar-Yosef and Francois R. Valla 521–56. Ann Arbor, Mich.
- Anderson-Gerfand, Patricia, Gérard Deraprahamian, and George Willcox. 1991. Les premières cultures de céréales sauvages et domestiques primitives au Proche-Orient Néolithique: Résultats préliminaires

- d'expériences à Jalès (Ardèche). *Cahiers de l'Euphrate* 5-6: 191-232.
- Aschmann, Homer. 1991. Human impact on the biota of Mediterranean-climate regions of Chile and California. In *Biogeography of Mediterranean invasions*, ed. F. H. Groves and F. di Castri, 33-41. Cambridge.
- Baker, H. G. 1970. Taxonomy and the biological species concept in cultivated plants. In *Genetic resources in plants*, ed. O. H. Frankel and E. Bennett 49-68. Oxford.
- Barker, Graeme. 1985. *Prehistoric farming in Europe*. Cambridge.
- Bar-Yosef, Ofer. 1986. The walls of Jericho: An alternative interpretation. *Current Anthropology* 27: 157-62.
1989. The PPNA in the Levant - An overview. *Paléorient* 15: 57-63.
- Bar-Yosef, Ofer, and Anna Belfer-Cohen. 1989. The origins of sedentism and farming communities in the Levant. *Journal of World Prehistory* 3: 447-97.
1992. From foraging to farming in the Mediterranean Levant. In *Transitions to agriculture in prehistory*, ed. Anne Birgitte Gebauer and T. Douglas Price 21-48. Madison, Wis.
- Bar-Yosef, Ofer, and Mordechai E. Kislev. 1986. Earliest domesticated barley in the Jordan Valley. *National Geographic Research* 2: 257.
- Bender, Barbara. 1978. Gatherer-hunter to farmer: A social perspective. *World Archaeology* 10: 204-22.
1981. Gatherer-hunter intensification. In *Economic archaeology*, ed. Alison Sheridan and Geoff Bailey. British Archaeological Reports, International Series 96. Oxford.
- Binford, Lewis R. 1968. Post-Pleistocene adaptations. In *New perspectives in archaeology*, ed. Sally R. Binford and Lewis R. Binford, 313-41. Chicago.
- Blumler, Mark A. 1992. Independent inventionism and recent genetic evidence of plant domestication. *Economic Botany* 46: 98-111.
- Blumler, Mark A., and Roger Byrne. 1991. The ecological genetics of domestication and the origins of agriculture. *Current Anthropology* 32: 23-54.
- Bohrer, Vorsila L. 1972. On the relation of harvest methods to early agriculture in the Near East. *Economic Botany* 26: 145-55.
- Braidwood, Robert J. 1960. The agricultural revolution. *Scientific American* 203: 131-46.
- Braidwood, Robert J., and Bruce Howe. 1960. *Prehistoric investigations in Iraqi Kurdistan*. Chicago.
- Brown, Terence A., Robin G. Allaby, Keri A. Brown, and Martin K. Jones. 1993. Biomolecular archaeology of wheat: Past, present and future. *World Archaeology* 25: 64-73.
- Burleigh, Richard. 1983. Appendix D: Additional radiocarbon dates for Jericho (with an assessment of all the dates obtained). In *Excavations at Jericho*, ed. Kathleen M. Kenyon and Thomas A. Holland, 760-5. London.
- Byrne, Roger. 1987. Climatic change and the origins of agriculture. In *Studies in the neolithic and urban revolutions*, ed. L. Manzanilla, 21-34. Oxford.
- Cauvin, Marie-Claire. 1974. Note préliminaire sur l'outillage lithique de la phase IV de Tell Mureybet (Syrie). *Annales Archéologiques Arabes Syriennes* 24: 58-63.
- Childe, V. Gordon. 1951. *Man makes himself*. 5th edition. New York.
1952. *New light on the most ancient East*. London.
- Close, Angela E., and Fred Wendorf. 1992. The beginnings of food production in the eastern Sahara. In *Transitions to agriculture in prehistory*, ed. Anne Birgitte Gebauer and Douglas Price, 63-72. Madison, Wis.
- Cohen, Mark N. 1977. *The food crisis in prehistory*. New Haven, Conn.
- COHMAP Members. 1988. Climatic changes of the last 18,000 years: Observations and model simulations. *Science* 241: 1043-52.
- Colledge, Susan. 1991. Investigations of plant remains preserved in epipalaeolithic sites in the Near East. In *The Natufian culture in the Levant*, ed. Ofer Bar-Yosef and Francois R. Valla, 391-8. Ann Arbor, Mich.
- Contenson, Henri de. 1985. La région de damas au néolithique. *Les Annales Archéologiques Arabes Syriennes* 35: 9-29.
- Contenson, Henri de, Marie-Claire Cauvin, Willem van Zeist, et al. 1979. Tell Aswad (Damascene). *Paléorient* 5: 153-76.
- Costantini, Lorenzo. 1981. The beginning of agriculture in the Kachi Plain: The evidence of Mehrgarh. In *South Asian archaeology 1981, proceedings of the 6th international conference of the Association of South Asian Archaeologists in Western Europe*, ed. Bridgit Allchin, 29-33. Cambridge.
- Crosby, Alfred W. 1986. *Ecological imperialism*. Cambridge.
- Davis, Simon J. M. 1987. *The archaeology of animals*. New Haven, Conn.
- Dennell, Robin W. 1983. *European economic prehistory*. New York.
- Feldman, Moche. 1976. Wheats. In *Evolution of crop plants*, ed. N. W. Simmonds, 120-8. London.
- Flannery, Kent V. 1965. The ecology of early food production in Mesoamerica. *Science* 147: 1247-56.
1969. Origins and ecological effects of early domestication in Iran and the Near East. In *The domestication and exploitation of plants and animals*, ed. Peter J. Ucko and Geoffrey W. Dimbleby, 73-100. London.
1973. The origins of agriculture. *Annual Review of Anthropology* 2: 271-310.
- Godelier, Maurice. 1970. *Sur les sociétés précapitalistes*. Paris.
- Graber, Robert Bates. 1992. Population pressure, agricultural origins, and global theory: Comment on McCorriston and Hole. *American Anthropologist* 94: 443-5.
- Grant, Verne. 1981. *Plant speciation*. New York.
- Harlan, Jack R. 1967. A wild wheat harvest in Turkey. *Archaeology* 20: 197-201.
1989. The tropical African cereals. In *Foraging and farming: The evolution of plant exploitation*, ed. David R. Harris and Gordon C. Hillman, 79-98. London.
- Harlan, Jack R., J. M. J. de Wet, and E. G. Price. 1973. Comparative evolution of cereals. *Evolution* 27: 311-25.
- Harlan, Jack R., and Daniel Zohary. 1966. Distribution of wild wheats and barley. *Science* 153: 1074-80.
- Harris, David R., V. M. Masson, Y. E. Berezkin, et al. 1993. Investigating early agriculture in Central Asia: New research at Jeitun, Turkmenistan. *Antiquity* 67: 324-38.
- Hayden, Brian. 1981. Research and development in the Stone Age: Technological transitions among hunter-gatherers. *Current Anthropology* 22: 519-48.
- Helbaek, Hans. 1964. Early Hassunan vegetables at es-Sawwan near Samarra. *Sumer* 20: 45-8.
1966. Pre-pottery neolithic farming at Beidha. *Palestine Exploration Quarterly* 98: 61-6.
- Henry, Donald O. 1981. An analysis of settlement patterns and adaptive strategies of the Natufian. In *Préhistoire du Levant*, ed. Jacques Cauvin and Paul Sanlaville, 421-32. Paris.

1985. Preagricultural sedentism: The Natufian example. In *Prehistoric hunter-gatherers: The emergence of cultural complexity*, ed. T. Douglas Price and James A. Brown, 365–84. New York.
1989. *From foraging to agriculture: The Levant at the end of the Ice Age*. Philadelphia, Pa.
- Heyne, E. G., and G. S. Smith. 1967. Wheat breeding. In *Wheat and wheat improvement*, ed. K. S. Quisenberry and L. P. Reitz, 269–306. Madison, Wis.
- Hillman, Gordon C. 1975. The plant remains from Tell Abu Hureyra: A preliminary report. In *The excavation of Tell Abu Hureyra in Syria: A preliminary report* (A. M. T. Moore). *Proceedings of the Prehistoric Society* 41: 70–3.
1989. Late palaeolithic plant foods from Wadi Kubbaniya in Upper Egypt: Dietary diversity, infant weaning, and seasonality in a riverine environment. In *Foraging and farming: The evolution of plant exploitation*, ed. David R. Harris and Gordon C. Hillman, 207–39. London.
1991. Comment on the ecological genetics of domestication and the origins of agriculture. *Current Anthropology* 32: 39–41.
- Hillman, Gordon C., Susan M. Colledge, and David R. Harris. 1989. Plant-food economy during the epi-palaeolithic period at Tell Abu Hureyra, Syria: Dietary diversity, seasonality, and modes of exploitation. In *Foraging and farming: The evolution of plant exploitation*, ed. David R. Harris and Gordon C. Hillman, 240–68. London.
- Hillman, Gordon C., and M. S. Davies. 1990. Measured domestication rates in crops of wild type wheats and barley and the archaeological implications. *Journal of World Prehistory* 4: 157–222.
- Hillman, Gordon C., Sue Wales, Frances McLaren, et al. 1993. Identifying problematic remains of ancient plant foods: A comparison of the role of chemical, histological and morphological criteria. *World Archaeology* 25: 94–121.
- Hodder, Ian. 1990. *The domestication of Europe*. Oxford.
- Hole, Frank A. 1984. A reassessment of the neolithic revolution. *Paléorient* 10: 49–60.
- Hole, Frank A., and Joy McCorriston. 1992. Reply to Graber. *American Anthropologist* 94: 445–6.
- Hopf, Maria. 1969. Plant remains and early farming in Jericho. In *The domestication and exploitation of plants and animals*, ed. Peter J. Ucko and Geoffrey W. Dimbleby, 355–9. London.
1983. Appendix B: Jericho plant remains. In *Excavations at Jericho*, Vol. 5, ed. Kathleen M. Kenyon and Thomas A. Holland, 576–621. London.
- Kenyon, Kathleen M. 1979. *Archaeology in the Holy Land*. Fourth edition. New York.
1981. *The architecture and stratigraphy of the Tell*. In *Excavations at Jericho*, Vol. 3, ed. Thomas A. Holland. London.
- Kimber, Gordon, and Moshe Feldman. 1987. *Wild wheat: An introduction*. Special Report No. 353, College of Agriculture, University of Missouri. Columbia.
- Kislev, Mordechai E. 1984. Emergence of wheat agriculture. *Paléorient* 10: 61–70.
- Kislev, Mordechai E., Ofer Bar-Yosef, and Avi Gopher. 1986. Early neolithic domesticated and wild barley from the Netiv Hagdud Region in the Jordan Valley. *Israel Journal of Botany* 35: 197–201.
- Ladizinsky, Gideon. 1985. Founder effect in crop plant evolution. *Economic Botany* 39: 191–9.
1989. Origin and domestication of the southwest Asian grain legumes. In *Foraging and farming: The evolution of plant exploitation*, ed. David R. Harris and Gordon C. Hillman, 374–89. London.
- Lamb, C. A. 1967. Physiology. In *Wheat and wheat improvement*, ed. K. S. Quisenberry and L. P. Reitz, 181–223. Madison, Wis.
- Le Houérou, Henri Noel. 1981. Impact of man and his animals on Mediterranean vegetation. In *Mediterranean-type shrublands*, ed. F. Di Castri, D. W. Goodall, and R. L. Specht, 479–522. Amsterdam.
- Lewis, Harlan. 1962. Catastrophic selection as a factor in speciation. *Evolution* 16: 257–71.
- Lilienfeld, F. A. 1951. H. Kihara: Genome-analysis in Triticum and Aegilops. X. Concluding Review. *Cytologia* 16: 101–23.
- Limbrey, Susan. 1990. Edaphic opportunism? A discussion of soil factors in relation to the beginnings of plant husbandry in South-West Asia. *World Archaeology* 22: 45–52.
- Löve, A. 1982. Generic evolution in the wheatgrasses. *Biologisches Zentralblatt* 101: 199–212.
- Mayr, Ernst. 1942. *Systematics and the origin of species*. New York.
- McCorriston, Joy. 1992. *The early development of agriculture in the ancient Near East: An ecological and evolutionary study*. Ph.D. dissertation, Yale University.
- McCorriston, Joy, and Frank A. Hole. 1991. The ecology of seasonal stress and the origins of agriculture in the Near East. *American Anthropologist* 93: 46–69.
- Moore, Andrew M. T. 1975. The excavation of Tell Abu Hureyra in Syria: A preliminary report. *Proceedings of the Prehistoric Society* 41: 50–77.
1979. A pre-neolithic farmer's village on the Euphrates. *Scientific American* 241: 62–70.
1985. The development of neolithic societies in the Near East. *Advances in World Archaeology* 4: 1–69.
1991. Abu Hureyra 1 and the antecedents of agriculture on the Middle Euphrates. In *The Natufian culture in the Levant*, ed. Ofer Bar-Yosef and Francois R. Valla, 277–94. Ann Arbor, Mich.
- Morris, R., and E. R. Sears. 1967. The cytogenetics of wheat and its relatives. In *Wheat and wheat improvement*, ed. K. S. Quisenberry and L. P. Reitz, 19–87. Madison, Wis.
- Nabhan, Gary P. 1989. *Enduring seeds*. San Francisco, Calif.
- Naveh, Zev. 1974. Effects of fire in the Mediterranean region. In *Fire and ecosystems*, ed. T. T. Kozłowski and C. E. Ahlgren, 401–34. New York.
- Naveh, Zev, and Joel Dan. 1973. The human degradation of Mediterranean landscapes in Israel. In *Mediterranean type ecosystems: Origin and structure*, ed. Francesco di Castri and Harold A. Mooney, 373–90. New York.
- Percival, John. 1921. *The wheat plant*. London.
- Purseglove, J. W. 1985. *Tropical crops: Monocotyledons*. Fifth edition. New York.
- Rindos, David. 1984. *The origins of agriculture*. New York.
- Rosenberg, Michael. 1990. The mother of invention: Evolutionary theory, territoriality, and the origins of agriculture. *American Anthropologist* 92: 399–415.
- Rosenberg, Michael, R. Mark Nesbitt, Richard W. Redding, and Thomas F. Strasser. 1995. Hallan Çemi Tepesi: Some preliminary observations concerning early neolithic subsistence behaviors in eastern Anatolia. *Anatolica* 21: 1–12.
- Runnels, Curtis, and T. H. van Andel. 1988. Trade and the origins of agriculture in the Eastern Mediterranean. *Journal of Mediterranean Archaeology* 1: 83–109.
- Schiemann, E. 1948. *Weizen, Roggen, Gerste. Systematik, Geschichte, und Verwendung*. Jena, Germany.
- Sherratt, Andrew. 1980. Water, soil and seasonality in early cereal cultivation. *World Archaeology* 11: 313–30.

- Sillen, Andrew. 1984. Dietary variability in the epipalaeolithic of the Levant: The Sr/Ca evidence. *Paléorient* 10: 79-84.
- Sillen, Andrew, and Julia A. Lee-Thorp. 1991. Dietary change in the late Natufian. In *The Natufian culture in the Levant*, ed. Ofer Bar-Yosef and François R. Valla, 399-410. Ann Arbor, Mich.
- Smith, Patricia, Ofer Bar-Yosef, and Andrew Sillen. 1984. Archeological and skeletal evidence for dietary change during the late Pleistocene/early Holocene in the Levant. In *Paleopathology at the origins of agriculture*, ed. Mark N. Cohen and George J. Armelagos, 101-30. New York.
- Tchernov, Eitan. 1991. Biological evidence for human sedentism in Southwest Asia during the Natufian. In *The Natufian culture in the Levant*, ed. Ofer Bar-Yosef and François R. Valla, Ann Arbor, Mich.
- Unger-Hamilton, Romana. 1989. The epi-palaeolithic southern Levant and the origins of cultivation. *Current Anthropology* 30: 88-103.
- van Zeist, Willem. 1970. The Oriental Institute excavations at Mureybit, Syria: Preliminary report on the 1965 campaign. Part III: The palaeobotany. *Journal of Near Eastern Studies* 29: 167-76.
1988. Some aspects of early neolithic plant husbandry in the Near East. *Anatolica* 15: 49-67.
- van Zeist, Willem, and Johanna Bakker-Heeres. 1982 (1985). Archaeobotanical studies in the Levant. I. Neolithic sites in the Damascus basin: Aswad, Ghoraifé, Ramad. *Palaeohistoria* 24: 165-256.
- 1984 (1986). Archaeobotanical studies in the Levant 3. Late-palaeolithic Mureybit. *Palaeohistoria* 26: 171-99.
- van Zeist, Willem, and Sytze Bottema. 1982. Vegetation history of the eastern Mediterranean and the Near East during the last 20,000 years. In *Palaeoclimates, palaeoenvironments and human communities in the eastern Mediterranean region in later prehistory*, ed. John L. Bintliff and Willem van Zeist, 277-321. 2 vols. Oxford.
- Vavilov, Nikolai I. 1951. The origin, variation, immunity, and breeding of cultivated plants. *Chronica Botanica* 13: 1-6.
- Watkins, Trevor, Douglas Baird, and Alison Betts. 1989. Qeremez Dere and the early aceramic neolithic in N. Iraq. *Paléorient* 15: 19-24.
- Watson, Patty Jo. 1991. Origins of food production in western Asia and eastern North America: A consideration of interdisciplinary research in anthropology and archaeology. In *Quaternary landscapes*, ed. Linda C. K. Shane and Edward J. Cushing, 1-37. Minneapolis, Minn.
- Wenke, Robert J. 1989. Egypt: Origins of complex societies. *Annual Review of Anthropology* 18: 129-55.
- Western, A. Cecilia. 1971. The ecological interpretation of ancient charcoals from Jericho. *Levant* 3: 31-40.
1983. Appendix F: Catalogue of identified charcoal samples. In *Excavations at Jericho*, vol. 5, ed. Kathleen M. Kenyon and Thomas A. Holland, 770-3. London.
- Wetterstrom, Wilma. 1993. Foraging and farming in Egypt: The transition from hunting and gathering to horticulture in the Nile Valley. In *The archaeology of Africa*, ed. Thurstan Shaw, Paul Sinclair, Bassey Andah, and Alex Okpoko, 165-225. London.
- Wilke, Philip J., Robert Bettinger, Thomas F. King, and James F. O'Connell. 1972. Harvest selection and domestication in seed plants. *Antiquity* 46: 203-9.
- Wright, Gary A. 1971. Origins of food production in southwest Asia: A survey of ideas. *Current Anthropology* 12: 447-77.
- Wright, Jr., Herbert E. 1977. Environmental change and the origin of agriculture in the Old and New Worlds. In *Origins of agriculture*, ed. C. A. Reed, 281-318. The Hague.
- Wright, Katherine I. 1994. Ground stone tools and hunter-gatherer subsistence in southwest Asia: Implications for the transition to farming. *American Antiquity* 59: 238-63.
- Zeven, A. C. 1980. The spread of bread wheat over the Old World since the neolithicum as indicated by its genotype for hybrid necrosis. *Journal d'Agriculture Traditionnelle et de Botanique Appliquée* 27: 19-53.
- Zohary, Daniel. 1970a. Centers of diversity and centers of origin. In *Genetic resources in plants*, ed. O. H. Frankel and E. Bennett, 33-42. Oxford.
- 1970b. Wild wheats. In *Genetic resources in plants*, ed. O. H. Frankel and E. Bennett, 239-47. Oxford.
1971. Origin of south-west Asiatic cereals: Wheats, barley, oats and rye. In *Plant life of southwest-Asia*, ed. Peter H. Davis, Peter C. Harper, and Ian C. Hedge, 235-60. Edinburgh.
1989. Domestication of the southwest Asia neolithic crop assemblage of cereals, pulses, and flax: The evidence from the living plants. In *Foraging and farming: The evolution of plant exploitation*, ed. David R. Harris and Gordon C. Hillman, 359-73. London.
- Zohary, Daniel, Jack R. Harlan, and A. Vardi. 1969. The wild diploid progenitors of wheat and their breeding values. *Euphytica* 18: 58-65.
- Zohary, Daniel, and Maria Hopf. [1988] 1993. *Domestication of plants in the Old World*. Second edition. Oxford.
- Zohary, Michael. 1983. Man and vegetation in the Middle East. In *Man's impact on vegetation*, ed. W. Holzner, M. J. A. Werger, and I. Ikusima, 163-78. The Hague.



II.B

Roots, Tubers, and Other Starchy Staples

II.B.1 🍌 Bananas and Plantains

Bananas represent one of the most important fruit crops, second only to grapes in the volume of world production (Purseglove 1988). J. F. Morton (1987) indicates that bananas are the fourth largest fruit crop after grapes, citrus fruits, and apples. Bananas and plantains are starchy berries produced by hybrids and/or sports of *Musa acuminata* Colla and *Musa balbisiana*. Rare genome contributions from another species may have occurred but are not yet well documented (Simmonds 1986). Additionally, *fe'i* bananas are obtained from *Musa troglodytarum*. Bananas may be differentiated from plantains on the basis of moisture content, with bananas generally averaging 83 percent moisture and plantains 65 percent (but intermediate examples may also be found) (Lessard 1992). Bananas may be eaten raw or cooked. Plantains are usually eaten cooked. Commonly, bananas which are eaten raw are referred to as dessert bananas. Throughout this essay, the term “bananas” is used to refer to both bananas and plantains.

Bananas, being primarily carbohydrates (22.2 to 31.2 percent), are low in fats, cholesterol, and sodium. Potassium levels are high (400 milligrams to 100 grams of pulp). Bananas are also good sources of ascorbic acid, 100 grams providing 13.3 to 26.7 percent of the U.S. RDA (Stover and Simmonds 1987). During ripening, the starch component is gradually converted to simple sugars (fructose, glucose, and sucrose), while the moisture content of the pulp increases. The time of conversion to simple sugars can also be used to differentiate plantains/cooking bananas (later conversion) from bananas that are eaten raw (earlier conversion).

Banana Plants

Bananas are monocarpic (fruiting once, then dying), perennial, giant herbs that usually are propagated via lateral shoots (suckers). Leaves are produced by a single apical meristem, which typically forms only a low

short stem or pseudobulb. The leaves are tightly rolled around each other, producing a pseudostem with a heart of young, emerging, rolled leaves ending with the terminal production of a huge inflorescence (usually sterile) and, finally, the starchy fruits: bananas or plantains.



Banana plant

Banana suckers emerge from axillary buds on the pseudobulb, thus providing a means of propagation as the fruits are commonly sterile. Suckers are either left in place as a part of a "mat," which includes the parent plant, or they may be removed for planting at new sites. Within a year after a sucker has been planted at a new site, the flowering stem will emerge at the apex of the pseudostem. The flowering stem will gradually bend over, producing a pendulous inflorescence (except in *fe'i* bananas, which have an erect inflorescence).

At the apical end of the stem are sterile male flowers, protected by large, often reddish, bracts (reduced or modified leaves). Higher up the stem are rows of biserially (in two series) arranged female (or hermaphroditic) flowers (Masefield et al. 1971). The banana fruits developing from the rows of flowers are commonly called "hands," with the individual fruits called "fingers" (Stover and Simmonds 1987). The entire inflorescence, having matured as fruit, may be called a "bunch."

Climate and Soil

Bananas are almost entirely restricted to largely tropical wet zones of the earth. Practically all banana cultivations fall within 30° latitude north and south of the equator (Simmonds 1966), with most of the large growing areas in the tropics between 20° north and south latitude. Bananas are very susceptible to cold temperatures and to drying environments. Their growth is limited by temperature in areas where water is not limited and by water availability in the warmest climates. A mean monthly temperature of 27° C is optimal, with temperatures below 21° C causing delayed growth (Purseglove 1988). Bananas are found growing under optimal conditions in wet or humid tropics when there are at least eight months per year with a minimum of 75 millimeters of rain per month (Stover and Simmonds 1987). Bananas also grow best under intense sunlight, with shading causing delayed growth, although the fruits may be sunburned, turning black, if exposed to excessive radiation.

Bananas will grow, and even produce fruit, under very poor conditions but will not produce an economically viable crop unless planted in relatively deep, well-drained soil (Morton 1987). Bananas can grow on loam, rocky sand, marl, volcanic ash, sandy clay, and even heavy clay, as long as water is not excessively retained in the soil matrix. Well drained, slightly acidic alluvial soils of river valleys offer optimal edaphic conditions.

General Uses

As already mentioned, banana and plantain fruits may be either cooked or eaten raw. The major usage of bananas is as a starch source for local consumption by tropical traditional cultures. Banana starch may be

consumed in a variety of products (see the section on *Cultural Uses*), with the bulk of the consumption consisting of very simple sugars mixed with fibers. The fruits represent significant exports from developing countries, particularly in the Neotropics. Bananas are very susceptible to damage during storage and shipping, which has certainly limited their value as foods imported into temperate, industrialized nations.

Where bananas are locally grown, the nonfruit parts of the plants are employed for a variety of purposes. Banana leaves are commonly used in addition to the fruits, with some varieties producing more desirable leaves than fruits. Fresh banana leaves serve as wrapping material for steamed or cooked foods and also as disposable meal platters. Fresh leaves are used medicinally in Rotuma, Samoa, and Fiji for the treatment of a variety of disorders, including headaches, menstrual cramps, and urinary tract infections. Young, unfolded leaves are employed as topical remedies for chest ailments, and the stem juice is used to treat gonorrhea (Uphof 1968).

Juice from fresh banana leaves occasionally serves as a light brown dye or stain. Because of their 30 to 40 percent tannin content, dried banana peels are used to blacken leather (Morton 1987). Dried leaves may be woven as house screens or be a source of fibrous strings for simple weaving or short-term structure construction. Dried leaves may also be employed to absorb salt water for transport to distant locations, where they are burned to produce a salty ash seasoning. Dried green plantains, ground fine and roasted, have reportedly served as a substitute for coffee (Morton 1987), and banana leaves have even been rolled as cigarette wrappers.

Of growing importance is the use of banana plants and fruits in livestock feed (Stover and Simmonds 1987; Purseglove 1988; Babatunde 1992; Cheeke 1992; Fomunyan 1992) and in shading or intercropping with yams, maize, cocoa, coconuts, areca nuts, and coffee (Stover and Simmonds 1987; Swennen 1990). Livestock are fed either dried and crushed fruits or fresh waste fruits and pseudostems with the leaves. Pigs, cattle, and rabbits have all been fed experimentally with mixtures of bananas and banana waste products. When used for intercropping or shading, bananas serve mainly as shade from intense sunlight for the crop of primary interest, but they also provide an intermediate crop while the farmer is waiting for production of his primary crop: cocoa, coconuts, areca nuts, or coffee.

Leaves of the related species *M. textilis* Nee have been used in and exported from the Philippines as a fiber source in the form of abaca. The abaca fiber is applied to the production of ropes, twines, hats, mats, hammocks, and other products requiring hard, strong fibers (Brown 1951; Purseglove 1988).

Biology

More than 500 varieties of bananas are recognized worldwide, although many of these are probably closely related lineages with differing regional names. Extensive research into the genetics, taxonomy, propagation, and distribution of bananas has been carried out by N. W. Simmonds (1957; 1962) and R. H. Stover (Stover and Simmonds 1987).

Taxonomy

Bananas and plantains have been taxonomically referenced by many different scientific names, including *Musa paradisiaca* L., *M. corniculata* Lour., *M. nana* Lour., and *M. sapientum* L. Each of these names is misleading, giving reference to a variety or group of varieties within a hybrid complex of extensively cultivated clones arising from *M. acuminata* Colla and *M. balbisiana* Colla. Simmonds (1962) suggested that the Latin binomials previously mentioned should all be abandoned and replaced by a designation of the clonal lineage represented by the ploidy (chromosome repetition number) and relative contributions of each of the diploid parent species.

The lineages are represented as groups of varieties/clones that share common ploidy levels and relative proportions of ancestral features (*M. acuminata* represented by "A" and *M. balbisiana* by "B"). Simmonds developed a system of scoring each variety on the basis of 15 characters, noting those characters present from the "A" and "B" parents. Additionally, many nonancestral (derived) somatic mutations have been identified, and these can be employed to differentiate the lineages of hybrids/autopolyploids (Stover and Simmonds 1987). It is possible that future systematic studies of bananas will use these ancestral and derived features as plesiomorphic (ancestral) and apomorphic (unique or derived) characters in cladistic and phenetic determinations of the relationships of banana cultivars.

Simmonds's designations are applied as follows for the common bananas with some examples of varieties within each group:

Group AA: Varieties 'Nino', 'Paka', 'Pisang lin', 'Sucrier', and 'Thousand Fingers'. These are primitive diploids found in Malesia, New Guinea, the Philippines, and East Africa.

Group AB: Varieties 'Ney Poovan' and some 'Lady's Fingers'. These diploid hybrids from India are of minor importance.

Group AAA: Varieties 'Cavendish', 'Dwarf Cavendish', 'Gros Michel', 'Highgate', and 'Robusta'. These varieties consist of African and Malesian triploids, which were important in the initial development of the commercial banana trade (mainly 'Gros Michel'), and remain important in much of the present trade ('Dwarf Cavendish').

Group AAB: Varieties 'Ae Ae', 'Apple', 'Brazilian', 'Giant Plantains', 'Hua Moa', 'Red holene', and 'Rhino horn'. These are triploids producing the lower moisture content fruits generally called plantains, which were initially developed in southern India.

Group ABB: Varieties 'ce Cream', 'Kru', 'Orinco', and 'Praying Hands'. These are triploids originating in India, the Philippines, and New Guinea. They are important staples in Southeast Asia, Samoa, and parts of Africa (Purseglove 1988).

Additionally, tetraploid (4 times the base chromosome number) hybrids AAAA, AB BB, AAAB, and AAB B have also been produced, but these are presently of little importance. They may, however, become important in the future (Purseglove 1988).

Fe'i bananas (*Musa troglodytarum* L.) differ from other common edible bananas in that they are diploids with erect inflorescences, have red-orange juice, and are not derived from hybrids of sports of *M. acuminata* or *M. balbisiana*. *Fe'i* bananas are prepared as plantains (cooked), having flesh which is rich orange-yellow to reddish in color.

Propagation

The clonal groups are propagated vegetatively with constant selection for desired traits (fruit elegance, flavor, and so forth, and resistance to diseases such as Panama disease, or to nematodes or corm borers). Of the 500 varieties of bananas that are recognized, about half are diploids, with most of the remainder being triploids (Purseglove 1988). Bananas take from 2 to 6 months or more to produce an inflorescence from a new sucker shoot.

Bananas can reproduce by seeds (in the wild primitive varieties) or, as is primarily the case, by suckers (in most cultivated and wild varieties). The suckers may be removed and planted in new locations to expand the range of the variety. Bananas moved from Malesia to Oceania, Africa, and Central America and the Caribbean by means of transplanted suckers.

Diseases

Bananas are afflicted with diseases caused by fungi, bacteria, and viruses. Fungal diseases include Sigatoka leaf spot, crown rot, anthracnose, pitting disease, brown spot, diamond spot, fusarial wilt (Panama disease), freckle disease, and rust. Bacterial infections include Moko, banana finger rot, and rhizome rot. "Bunchy top" is the only widespread virus that attacks bananas. It is transmitted by an aphid vector and can be controlled with insecticides (Stover and Simmonds 1987).

Bananas are also susceptible to damage from nematodes, insect larvae, and adult insects. Caterpillar defoliators are common but do not usually cause

sufficient destruction to impact production levels. Boring worms cause some damage to the pseudostems and rhizomes, but nematodes account for most of the damage to those tissues. Thrips and beetles chew and suck on the pseudostems, fruits, and suckers, leaving unsightly scars that reduce the market value of the fruits. Nematodes can cause significant damage, particularly to the commercial 'Cavendish' varieties. The nematodes will commonly attack plants that have been weakened by other pathogens, such as Sigatoka leaf spot, which promotes rotting of the pseudostem, rhizome, and roots.

Of the pathogens listed, fusarial wilt (Panama disease) and Sigatoka leaf spot have had the greatest impact on both the local and commercial production of bananas. Fusarial wilt wrought devastating losses of bananas in the Neotropics between 1910 and 1955, causing commercial banana producers to switch from the 'Gros Michel' variety to 'Cavendish' varieties, which are more resistant to fusarial wilt. Sigatoka leaf spot involves three closely related fungi that destroy the banana leaves and decrease transportability of the fruits (Stover and Simmonds 1987).

History

The wild ancestors of edible bananas (*M. acuminata* Colla and *M. balbisiana* Colla), except for the *fe'i* bananas, are centered in Malesia, a term that refers to the entire region from Thailand to New Guinea – roughly the main trading area of the Malay mariners. Simmonds (1962) has indicated that probably crosses and/or autopolyploidy of these wild ancestors originally took place in Indochina or the Malay Archipelago. Subspecies of *M. acuminata* have also been transported to locations as distant as Pemba, Hawaii, and Samoa, where they may have contributed to the production of new varieties.

The subspecies *malaccensis* produces edible diploid fruits via parthenocarpy and female sterility. These characters would have been fostered by human selection and vegetative propagation, transforming bananas from jungle weeds into a productive crop (Purseglove 1988). According to Simmonds (1962):

Edibility arose in subsp. *malaccensis* near the western edge of the range of *M. acuminata*, and, perhaps, in other subspecies independently; male-fertile edible clones were carried to other areas and intercrossed and outcrossed to local wild forms, generating new phenotypes and new structural chromosome combinations in the process; selection retained the best and the most sterile of these, which, under prolonged clonal propagation, accumulated still more structural changes until, finally, total sterility supervened and clonal propagation became obligatory.

Distribution

A Burmese legend relates that humans first realized bananas could be eaten when they observed birds eating them. The common Burmese generic name for bananas is *bnget pyaw*, meaning "the birds told" (Simmonds 1957; Lessard 1992). Fijians tell a story about a young girl whose lover disappeared while holding her hands in farewell. He was replaced by a banana sucker, which grew "hands" of banana "fingers" that represented the outstretched hands of the lost lover (Reed and Hames 1993).

Bananas are thought to have been distributed from western Melanesia into eastern Melanesia, Polynesia, and Micronesia during the time of aboriginal migrations into these areas. Linguistic evidence indicates a common center of origin of some Polynesian or Micronesian banana varieties in Indo-Malaysia. The movements of these varieties can be traced through two dispersals. The first involved movement into the Philippines, then into Micronesia and, eventually, Polynesia. The second dispersal involved movement into Melanesia first, with a secondary dispersal into parts of Polynesia (Guppy 1906). In the later dispersal, seeded varieties and the *fe'i* bananas constituted the initial wave of introduction, followed by successive waves of nonseeded varieties imported from west to east, penetrating into Oceania such that more varieties may be found in western, and fewer in eastern, Oceania. (This is a general trend for the common banana varieties but not for the *fe'i* bananas, which have the greatest diversity in the extreme east of the distribution.)

Bananas may initially have been introduced into Africa by Arab traders who brought the plants from Malaysia. But they may also have arrived earlier with Indonesians who brought the fruit to Madagascar. Or, of course, they could have been introduced even earlier by unknown individuals from unknown sources. Regardless, the plants subsequently spread across tropical Africa from east to west. Simmonds (1957) indicates that the banana entered the Koran as the "tree of paradise." The generic name *Musa* is derived from the Arabic word *mouz*, meaning banana (Purseglove 1988). Linnaeus applied the name *Musa paradisiaca* as a reference to this ancient terminology.

The name "banana" came from the Guinea coast of West Africa and was introduced along with the plant into the Canary Islands by the Portuguese (Purseglove 1988). At least one clone was taken from the Canary Islands to Hispaniola in 1516. Bananas were carried as food for slaves who originated in areas of traditional banana cultivation, and J. W. Purseglove (1988) surmises that 1516 may mark the first introduction of bananas into the Caribbean and tropical America. Alternatively, bananas may have arrived in the Neotropics via the Spanish trade route from the Philippines. The Spaniards used the name *plátano* for bananas, from which the term "plantain" has been derived (Purseglove 1988).

The Economic Importance of Bananas

The secondary centers of banana distribution in Africa, Central America, and the Caribbean have become the greatest consumers and exporters of bananas. Estimates of the total banana production of the world range from 20 to 40 million tons (Simmonds 1966; Stover and Simmonds 1987), with 4 to 5 million tons per year entering international trade (Morton 1987; Purseglove 1988: 376). Africa is the largest producer of bananas, with some sources saying that half of the world's bananas are produced there. But most African bananas are consumed locally, although some are exported to Europe. Ecuador (the world's largest banana exporter), Colombia, Costa Rica, Honduras, Jamaica, and Panama all export bananas to the United States and Europe, whereas the Philippines and Taiwan export bananas to other Asian countries, particularly Japan.

Three-fourths of internationally traded bananas are grown in Central and South America and the Caribbean, with much of this trade controlled by the United Fruit Company and the Standard Fruit Company. The former has enormous land concessions, regional shipping interests, and distribution networks within the United States (Purseglove 1988). United Fruit developed its economic empire beginning in 1874 in Costa Rica, and subsequently expanded throughout the region. The bananas produced were all of the 'Gros Michel' variety until 1947, when because of increasing losses from Panama disease, the 'Robusta' and other disease-resistant varieties began to replace the 'Gros Michel.' Panama disease (causing leaves to wilt and die) and Sigatoka disease (causing decay spots on leaves and either death of the plant or a greatly reduced crop output) are the two major diseases that have limited the international economic potential of bananas and have also driven the study and introduction of various disease-resistant varieties of the fruit. The economic value of bananas in international trade is, however, a secondary consideration; their major importance is in providing basic local nutrition for many rural populations in tropical, Third World countries.

It is interesting to note that both the *fe'i* bananas and the common bananas have attained their greatest agricultural and human importance in areas far removed from their centers of origin. *Fe'i* bananas originated in Melanesia, spreading and diversifying into the central Pacific, and the common bananas originated in Malesia, spreading and diversifying in Africa, Oceania, India, and, most recently, the Neotropics.

Specific Cultural Usages

Southeast Asia

The origin of bananas was probably within the cultures that developed in Malaysia, Indonesia, Thailand, and Burma. These cultures have long traditions of banana usage. Mature fruits are commonly eaten, but

in addition, young fruits are pickled and male buds are consumed as a vegetable in Malaysia and Thailand. Sap from the variety 'Pisang klutum' is mixed with soot and used to give a black color to bamboo basket-work in Java. Also in Java, the fruit stems surrounded by the leaf sheaths of the 'Pisang baia' variety are employed in the preparation of a type of sweetmeat (Uphof 1968). Flowers may be removed from the buds and used in curries in Malaysia. Ashes from burned leaves and pseudostems serve as salt in the seasoning of vegetable curries. Banana plants may be placed in the corners of rice fields as protective charms, and Malay women may bathe with a decoction of banana leaves for 15 days after parturition (Morton 1987).

Although Southeast Asia/Malesia is considered to be the origin of the distribution of bananas, it is important to note that bananas never became as important there as they did in parts of Africa, Oceania, and, more recently, the Neotropics, although production of bananas is increasing in Southeast Asia (Morton 1987; Sadik 1988). Such a relative lack of importance is perhaps connected with the presence of two other competing Southeast Asian starch crops: rice and sago. Rice offers a storable starch source with potentially greater stability of production and yield than bananas. Sago palms of the genus *Metroxylon* are present throughout Malesia in the same areas in which bananas may be found. But as sago production drops off and disappears, there is increased dependence upon bananas and increased diversity in some varieties, including the *fe'i* bananas.

Two areas that certainly must have received some of the earliest diffusions of bananas out of Malesia are the Philippines and India. In these areas, banana cultivation became much more important and variations in usage evolved. Filipinos eat not only the cooked or raw fruits of fresh banana but also the flowers. They employ the leaves and sap medicinally and extract fibers from the leaves. Young inflorescences are eaten both boiled as a vegetable and raw in salads (Brown 1951).

Banana fibers are used in the production of ropes and other products that require durability and resistance to saltwater. Fibers from the sheathing leafstalks are employed in the manufacture of a light, transparent cloth known as "agna" (Brown 1951). Wild banana leaves are used extensively as lining for cooking pots and earthen ovens and for wrapping items that are to be sold in markets. The Filipinos have effectively developed the banana as an export crop with up to a half million tons sent yearly to Japan (Morton 1987).

In India, "the banana plant because of its continuous reproduction is regarded by Hindus as a symbol of fertility and prosperity and the leaves and fruits are deposited on doorsteps of houses where marriages are taking place" (Morton 1987). Virtually the entire above-ground portion of the banana plant is eaten in India. The fruits are cooked or eaten raw, with no

clear distinction between plantains and bananas. The young flowers are eaten raw or cooked. The pseudostem may be cooked and eaten as a vegetable or may be candied with citric acid and potassium metabisulphite. India is currently the leading producer of bananas in Asia, with virtually the entire crop employed for domestic purposes (Morton 1987).

Africa

In Africa, bananas reach their greatest importance as a starchy food (Purseglove 1988). Throughout African regions where bananas grow, 60 million people (34 percent of the population) derive more than 25 percent of their calories from plantains and bananas (Wilson 1986). Within Africa, many tropical traditional cultures have come to depend heavily upon bananas as a starch source (Sadik 1988). For example, the Buganda in Uganda typically consume 4 to 4.5 kilograms of bananas per person daily. Tanzanians and Ugandans produce large quantities of beer from bananas ripened evenly in pits. These bananas, after becoming partially fermented, are trampled to extract the juice, which is then mixed with sorghum flour and old beer and allowed to ferment for 12 or more hours. The beer is drunk by people of all ages and has become an important part of the diet (Purseglove 1988).

Sweetmeats, made from dried banana slices, serve as famine foods, preserves, and desserts. Flour can be produced from ripe or unripe fruits, and the flowers are employed as an ingredient in confections. Banana flour is sometimes called Guiana arrowroot, a reference to its importance in West Africa (Uphof 1968). The fruits can be used medicinally for children who are intolerant of more complex carbohydrates and for adults with various intestinal complaints (Purseglove 1988). Additionally, banana pseudostem fibers are used as fishing line in parts of West Africa (Morton 1987).

In West Africa, as in many other parts of the world, bananas are grown in compound gardens and in and around village backyards (Swennen 1990). Proximity to the human population allows for harvesting of the unevenly ripening bananas, a few at a time from each bunch, and the human presence wards off birds and other animals that eat the fruits. In this situation, banana plants typically grow on local refuse dumps, which become rich in nutrients from decaying food products. Banana plants growing in rich soils tend to produce larger bunches, which may become so heavy that the plant falls over. Traditional farmers will place one or more prop poles under the base of growing bunches in order to keep the plants upright.

Oceania

Fried or steamed bananas are staples in many Polynesian cultures. Rotumans serve fried bananas as part of meals containing other dishes; alternatively, an entire

meal may consist of fried bananas. Banana fruits are frequently grated or pounded, mixed with coconut cream and fruit juices, and served as a thick beverage. Banana bunches are often found hanging at the edges of cookhouses and ceremonial structures in which a celebration is occurring. Hung in a cookhouse, bananas will slowly ripen over a period of days or weeks, allowing for gradual usage of the bunch. Bananas hung for ceremonial feasts and celebrations will be pit ripened (see below) in advance so that the entire bunch may be eaten ripe on the same day.

Many bananas ripen unevenly, with the more basal (usually higher on a pendulous raceme) fruits ripening first. In these cases, the basal bananas will have ripened and fallen off or been eaten before the apical fingers have even begun to ripen. This natural tendency makes it difficult to cut and use an entire bunch of bananas at one time. A traditional method, which is often used in Polynesia (and elsewhere) to promote even ripening and to increase the rate of ripening, is as follows: A pit (approximately 1 to 2 meters deep) is dug in the ground. It is made sufficiently large to hold several bunches of bananas (up to 10 to 15). The pit is lined with leaves, and fires are built in the edges of the pit. Unripened, but mature, banana bunches are cut from the pseudostems and placed in the pit. The bananas are covered with leaves and the fires are stoked to burn slowly. The pit is covered with soil and left for 3 to 7 days. This process both heats the bananas and extracts oxygen from within the pit. When the pit is opened, the bunches of bananas are found to be entirely and uniformly ripened. Perhaps this ripening process occurs because of increased concentrations of the ethylene produced by the earliest ripening bananas. This in turn would speed the ripening of neighboring fingers, and the exclusion of oxygen and insects would prohibit deterioration of the fruits that ripened first (Stover and Simmonds 1987). The speed of the process is increased by the heat generated by the slowly smoldering fires.

The Neotropics

Neotropical tribes, such as the Waimiri Atroari of Brazilian Amazonia, have incorporated bananas into their diets as both food and beverage. Other neotropical traditional cultures have incorporated bananas not only as food but also as medicinals for stomach ulcers, as antiseptics, and as antidiarrheal remedies (Milliken et al. 1992).

As already mentioned, three-quarters of the worldwide production of bananas for international trade is produced in Central and South America and the Caribbean. This trade is extremely important to the local economies in Colombia, Costa Rica, Honduras, Jamaica, and Panama. The significance of bananas in the United States and Europe has been entirely a function of production in the Neotropics.

The United States

Bananas commonly obtained in the United States may be 'Dwarf Cavendish' or 'Gros Michel' varieties or clones closely related to the 'Dwarf Cavendish' (Simmonds 1986). About 12 percent of the world production of bananas involves these AAA cultivars (Stover and Simmonds 1987). The bananas eaten in temperate countries are typically picked prior to ripening of any of the bananas in a bunch. The bunches are broken down into individual "hands" for shipping from the tropics. Bananas are shipped - usually within 24 hours of harvest - in containers that maintain low storage temperatures from 12° C to -25° C. The unripe bananas may be stored for up to 40 days under controlled conditions of cool temperatures and ethylene-free environments. When ripening is desired, the temperature is raised to 14 to 18° C and ethylene gas is sprayed on the bananas, resulting in rapid, uniform ripening in 4 to 8 days (Stover and Simmonds 1987).

Will C. McClatchey

Bibliography

- Babatunde, G. M. 1992. Availability of banana and plantain products for animal feeding. In *Roots, tubers, plantains and bananas in animal feeding*, ed. D. Machin and S. Nyvold, 251-76. Rome. FAO Animal Production and Health Paper No. 95.
- Brown, W. H. 1951. *Useful plants of the Philippines*, Vol. 1. Manila.
- Cheeke, P. R. 1992. Feeding systems for tropical rabbit production emphasizing roots, tubers and bananas. In *Roots, tubers, plantains and bananas in animal feeding*, ed. D. Machin and S. Nyvold, 235-50. Rome. FAO Animal Production and Health Paper No. 95.
- Fomunyan, R. T. 1992. Economic aspects of banana and plantain use in animal feeding: The Cameroon experience. In *Roots, tubers, plantains and bananas in animal feeding*, ed. D. Machin and S. Nyvold, 277-89. Rome. FAO Animal Production and Health Paper No. 95.
- Guppy, H. B. 1906. Observations of a naturalist in the Pacific between 1896 and 1899. London.
- Lessard, W. O. 1992. *The complete book of bananas*. Homestead, Fla.
- Masefield, G. B., M. Wallis, S. G. Harrison, and B. E. Nicholson. 1971. *The Oxford book of food plants*. London.
- Milliken, W., R. P. Miller, S. R. Pollard, and E. V. Wandelli. 1992. *The ethnobotany of the Waimiri Atroari Indians of Brazil*. Kew, England.
- Morton, J. F. 1987. *Fruits of warm climates*. Greensboro, N.C.
- Purseglove, J. W. 1988. *Tropical crops: Monocotyledons*. Essex, England.
- Reed, A. W., and I. Hames. 1993. *Myths and legends of Fiji and Rotuma*. Auckland.
- Sadik, S. 1988. *Root and tuber crops, plantains and bananas in developing countries*. FAO Plant Production and Protection Paper No. 87. Rome.
- Simmonds, N. W. 1957. *Bananas*. London.
1962. *The evolution of bananas*. London.
1966. *Bananas*. Second edition. London.
- Bananas, *Musa cvs.* In *Breeding for durable resistance in perennial crops*, ed. N. W. Simmonds, 17-24. Rome. FAO Plant Production and Protection Paper No. 70.
- Stover, R. H., and N. W. Simmonds. 1987. *Bananas*. Third edition. London.
- Swennen, R. 1990. *Plantain cultivation under West African conditions*. Ibadan, Nigeria.
- Uphof, J. C. Th. 1968. *Dictionary of economic plants*. New York.
- Wilson, G. F. 1986. Status of bananas and plantains in West Africa. In *Banana and plantain breeding strategies*, ed. G. J. Persley and E. A. DeLanghe, 29-35. Cairns, Australia.

II.B.2 🐉 Manioc

A tropical root crop, manioc is also known as cassava, *mandioca*, *aipim*, the tapioca plant, and *yuca*. The term cassava comes from the Arawak word *kasabi*, whereas the Caribs called the plant *yuca* (Jones 1959). The word manioc, however, is from *maniot* in the Tupí language of coastal Brazil; *mandioca* derives from *Mani-ôca*, or the house of Mani, the Indian woman from whose body grew the manioc plant, according to Indian legends collected in Brazil (Casudo 1984). Domesticated in Brazil before 1500, *Manihot esculenta* (Crantz), formerly termed *Manihot utilisima*, is a member of the spurge family (Euphorbiaceae), which includes the rubber bean and the castor bean (Cock 1985).

The manioc plant is a perennial woody shrub that reaches 5 to 12 feet in height, with leaves of 5 to 7 lobes that grow toward the end of the branches. The leaves are edible and may be cooked like spinach, but in terms of food, the most significant part of the plant is its starchy roots, which often reach 1 to 2 feet in length and 2 to 6 inches in diameter. Several roots radiate like spokes in a wheel from the stem, and each plant may yield up to 8 kilograms of roots (Jones 1959; Cock 1985; Toussaint-Samat 1992).

There are two principal varieties of manioc - the sweet and the bitter. The sweet varieties have a shorter growing season, can be harvested in 6 to 9 months, and then can simply be peeled and eaten as a vegetable without further processing. If not harvested soon after maturity, however, sweet manioc deteriorates rapidly. The bitter varieties require 12 to 18 months to mature but will not spoil if left unharvested for several months. Thus, people can harvest them at their leisure. The main disadvantage to the bitter varieties is that they may contain high levels of cyanogenic glycosides, which can cause prussic-acid poisoning if the roots are not processed properly (Jones 1959; Johns 1990).

An obvious question is that given the threat of poisoning, why would Amerindians have domesticated

such a plant? The answer lies in its many advantages. It is a crop that does well in the lowland tropics where there is a warm, moist climate and no frost, although there are "cold-tolerant varieties" of the plant in the Andes (Cock 1985). In addition, manioc yields good results on soils of low fertility, and it will also tolerate acidic soils more readily than other food staples. One of the most important characteristics of manioc, however, is its ability to survive natural disasters, such as droughts. When other food crops dry up, people survive on manioc roots. Similarly, where storms frequently sweep the land, high winds do not kill the roots, even if they damage the foliage. New shoots soon form, while the roots continue to nourish people and prevent starvation.

Manioc roots are also resistant to locust plagues (an important consideration in Africa) and to destructive predators, such as wild pigs, baboons, and porcupines (Johns 1990). Once processed, manioc can be preserved and stored in a tropical climate as *farinha* (manioc meal) or as a bread (*pan de tierra caliente*, as it was called by late colonial Mexicans [Humboldt 1811]). To produce more manioc plants, farmers do not have to set aside edible roots; instead, they use stem cuttings or seeds to propagate the plant (Cock 1985).

As a food, manioc is very versatile because it can be boiled in a mush, roasted, baked, and even consumed as a pudding (tapioca) or alcoholic beverage (Aguilar 1982). When fresh, the manioc root is primarily a starch, a source of carbohydrates. But the leaf has protein and vitamin A, and the fresh roots may contain calcium, vitamin C, thiamine, riboflavin, and niacin. However, the nutritional value of the roots varies with processing, as vitamins may be leached, and even destroyed, when they are soaked and boiled (Jones 1959). Thus, as a rule, manioc must be supplemented with other foodstuffs in order for a population to avoid malnutrition. In many parts of the world, especially Asia, it also serves as an animal feed.

William Jones (1959: 29) has observed that modern methods for processing manioc roots derive from Indian methods. In order to consume the bitter varieties, they had to detoxify the plant by grating and soaking it to remove the toxic chemicals (Johns 1990). To prepare the coarse meal, known as *farinha de mandioca* (also *farinha de pau*) in Brazil, women, who traditionally process manioc in Amerindian societies, have to wash, peel, and scrape the roots. Some prehistoric populations in South America and the Caribbean even used their upper front teeth in processing manioc.

Using a flat piece of wood studded with small pointed stones as a grater, women convert the roots into a snowy white mass, which is then placed in a *tipiti*, a long cylindrical basket press similar to a Chinese "finger trap." The two ends of the *tipiti* are pulled apart, with one end tied to the ground and the other to the branch of a tree. After the excess liquid has

been squeezed out, the pulpy mass is removed, put through a sieve, and then placed on a flat ceramic griddle or metal basin where it is toasted over a low fire. The *farinha* can be kept for months and then eaten dry or mixed with water as a gruel (Jones 1959; de Léry 1990; Toussaint-Samat 1992; and personal observation, Tikuna village, Peru, 1975).

Origins

Although scholars agree that manioc was domesticated in the Americas, there is doubt about the exact location, even though the largest variety of species survive in Brazil. Possible areas of origin include Central America, the Amazon region, and the northeast of Brazil. Milton de Albuquerque (1969), a specialist on manioc in Amazonia, reported that the most primitive form of the plant is found in central Brazil in the state of Goiás, a region subject to prolonged dry seasons, but he believes that the backlands of the state of Bahia are its most probable point of origin. The oldest archaeological records in Brazil, however, come from the Amazon region, where ceramic griddles used in manioc preparation have been found in pre-Columbian sites (Lathrap 1970; Roosevelt 1980).

Of much greater antiquity, however, are remains of the manioc plant that have been discovered in South American excavations in and near the Casma Valley of northern Peru. These have been dated to 1785 B.C. (Langdon 1988). In Mexico, cassava leaves that are 2,500 years old have been found, along with cassava starch in human coprolites that are 2,100 to 2,800 years old (Cock 1985). Preclassic Pacific coast archaeological sites in Mesoamerica have yielded evidence of manioc, which was a staple of the Mayan civilization (Tejada 1979; Morley and Brainerd 1983). The quality of the evidence for manioc in Mesoamerica has led some authors to believe that manioc was first domesticated in Central America rather than Brazil, or that there may have been two regions of origin. Another possibility might be that bitter manioc was domesticated in northern South America, whereas sweet cassava was domesticated in Central America (Cock 1985).

Archaeological evidence for ancient manioc usage also exists in the Caribbean. According to Suzanne Levin (1983: 336), manioc griddles have been found in archaeological excavations in the Lesser Antilles on islands such as St. Kitts, St. Vincent, Antigua, and Martinique. Both the Arawaks and Caribs utilized them, and as they migrated to the islands from South America, they undoubtedly carried with them manioc and the technological knowledge necessary to propagate, cultivate, and process it.

When the Spanish reached the Caribbean and Central America, they discovered the indigenous populations cultivating manioc, a plant they termed *yuca*. Thus, the earliest European description of manioc dates from 1494. In describing the first voyage of

Columbus, Peter Martyr referred to “venomous roots” used in preparing breads (Pynaert 1951). The Portuguese encountered manioc after 1500 on the coast of Brazil. Other sixteenth-century observers, such as the German Hans Staden [1557] (1974) and the Frenchman Jean de Léry [1578] (1990), have left valuable descriptions of what the Brazilians term the most Brazilian of all economic plants because of its close links to the historical evolution of Brazil (Aguiar 1982).

As the Portuguese divided Brazil into captaincies and contested the French for its control, they employed their slaves, both Indian and African, in cultivating food crops, including manioc. Unlike other areas of the Americas, where Europeans refused to adopt Amerindian crops such as amaranth, in sixteenth-century Brazil manioc rapidly became the principal food staple of coastal settlers and their slaves. As the most fertile lands in Pernambuco and the Recôncavo of Bahia were converted to sugar plantations, less prosperous farmers grew manioc on more marginal lands for sale to planters and to people in nearby towns.

When the Dutch invaded Brazil in the seventeenth century, they mastered the Brazilian system of large-scale sugar cultivation, as well as the Brazilian system of food production for plantation slaves, meaning essentially manioc cultivation. After the Dutch were expelled from Brazil in 1654, they carried the Brazilian system to the Caribbean and, henceforth, West Indian planters, such as those on Martinique, “obliged” African slaves to cultivate food crops, including manioc, on their provision grounds (Tomich 1991). Thus, manioc became a part of the slave diet in the Caribbean as in Brazil.

Manioc also enabled fugitive slaves living in maroon communities or in *quilombos* in Brazil to survive on marginal lands in remote or difficult terrain. Although descriptions of manioc cultivation in *quilombos* have usually surfaced in Brazilian records only upon the destruction of *quilombos*, such as Palmares (*Relação das guerras feitas aos Palmares* . . . [1675–1678] 1988), Richard Price (1991) has been able to document the cultivation of manioc over two hundred years by the Saramaka maroons of Suriname and their descendants. They raised it in the 1770s as well as in the 1970s (Price 1991).

Manioc, which was closely linked to plantation slavery in the Americas, was also the food staple that enabled the conquest of the tropics. In the pre-Columbian period, a bread made from manioc permitted long-distance trade and exploration in South America as well as lengthy war expeditions. During the European conquest, the Spanish and Portuguese forces in the tropics rapidly adopted manioc bread. Later, rations were doled out to troops fighting the frontier wars of tropical Latin America, and even in twentieth-century Brazil, military forces received “farinha de guerra” (war meal) to sustain them in their

garrisons (Cascudo 1984). The *Bandeirantes* from São Paulo, who explored Brazil’s vast interior and discovered gold in the late seventeenth century, were able to do so because of manioc. The Guaraní Indians they pursued and often enslaved in Paraguay also raised manioc (as their ancestors had done for millennia) in the Jesuit missions that gave them refuge (Reff 1998).

In addition to slaves, maroons, soldiers, and explorers, free peasants also subsisted on manioc in Brazil, existing on the margins of plantation society and in the interior. In cultivating manioc for their own subsistence, they frequently produced a surplus, which they sold to planters and to townspeople. Often black and mulatto dependents (*agregados*) of the great sugar planters, these peasants escaped slavery by raising manioc and marketing it (Karasch 1986). Manioc thus supported the emergence of a free peasantry in the shadow of the Latin American plantation societies.

By the end of the colonial period, manioc had emerged as the principal food staple of the enslaved and impoverished in tropical Latin America and, as such, its cultivation, transportation, and commerce contributed greatly to the internal economy of Latin America. Unfortunately for the Latin Americans, however, manioc did not find a niche in global trade because the Portuguese had long before introduced the plant to the rest of the tropical world.

Africa

From Brazil, the Portuguese carried manioc to their stations along the Upper Guinea coast in West Africa and to the Kingdom of Kongo in northern Angola. Although manioc was not readily adopted in sixteenth-century West Africa, it was successfully introduced into the Kingdom of Kongo in what is now the modern country of Angola. Jones (1959) attributes the success of manioc in central Africa to the close ties between the Portuguese and the BaKongo beginning in the 1480s. An oral tradition with reference to the first Portuguese on the Congo-Angola coastline in the 1480s describes the arrival of “white men” who “brought maize and cassava and groundnuts and tobacco” (Hall 1991: 169).

The first documented reference to manioc comes a century later in 1593 in a letter from Sir Richard Hawkins regarding the seizure of a Portuguese ship engaged in the slave trade. Its cargo was, he reported, “meale of cassavi, which the Portingals call *Farina de Paw* [sic]. It serveth for marchandize in Angola, for the Portingals foode in the ship, and to nourish the negroes which they should carry to the River of Plate” (Jones 1959: 62). Thus, by the late sixteenth century, manioc meal was already an item of trade in Angola, as well as a food for the slaves in transit to the Americas. It would serve as a staple of the slave trade until that trade’s effective abolition in the mid-nineteenth century.

By the 1660s, manioc was an important food in northern Angola, according to a pair of Europeans who visited Luanda. Oral traditions from nearby areas stress the borrowing of manioc techniques from the Kingdom of Kongo; its use as a vegetable (the sweet variety?) before the people learned more complex processing techniques; and the crop's resistance to locusts. As Jones notes, the Africans who most enthusiastically adopted manioc were those living in the tropical rain forests of the Congo Basin, rather than the people of the grasslands, where maize, millet, or sorghum were cultivated.

In the Americas, the culture of manioc had been perfected by forest peoples, and in Africa, too, it was the forest peoples who welcomed this addition to their food supply. Clearly, it was an important addition, and more than 300 years later, the people of Zaire and Congo continued to consume manioc at an average of about 1 kilogram per person per day, which would have supplied over 1,000 calories. They also utilized cassava leaves as a source of vegetable protein in sauces and soups (Cock 1985: 10).

In West Africa, however, the cultivation of manioc spread more slowly. First introduced from Brazil to Portuguese Guiné and the island of São Tomé, the root had become an important food crop on São Tomé and the islands of Príncipe and Fernando Pó by 1700, but it found little acceptance on the mainland until the nineteenth century. At the end of the seventeenth century, William Bosman, a Dutch factor at El Mina (now Ghana), identified the foodstuffs of Liberia, Ghana, Dahomey, and Benin as yams and maize. He also listed millet, rice, sweet potatoes, beans, and groundnuts. Apparently, the Africans had accepted the American crops of corn, beans, and groundnuts but not manioc. As Jones (1959) notes, this may have been because farmers could treat corn as another cereal crop, and sweet potatoes were similar to yams. But another reason is that the complex processing methods required by manioc could only have been mastered through close ties to people familiar with them, and the Portuguese did not have the same kind of colonial relationship with the West African states that they did with Angola.

Extensive manioc cultivation did develop in the early nineteenth century, however, as former slaves returned from Brazil to West Africa. In 1910, for example, a Dahoman chief reported the oral tradition that a returned Brazilian (Francisco Felix da Souza) "had taught the Dahomans how to prepare manioc so they could eat it without becoming ill." Returned ex-slaves from Brazil in the 1840s also spread manioc cultivation in western Nigeria (Ohadike 1981: 389). Thus, Africans who had learned to process manioc into *farinha* while enslaved in Brazil played a key role in instructing those in West Africa how to utilize it. There it became known as *gari* or *garri* (Jones 1959).

The further diffusion of manioc in West Africa in the nineteenth and twentieth centuries was linked to

European colonialism. As European labor demands disrupted traditional systems of food production, the colonialists sought new food crops to ward off hunger and famine. Migratory workers also dispersed knowledge of manioc cultivation inland from the coast. D. C. Ohadike (1981) argues that the influenza pandemic of 1918 that disrupted traditional agriculture based on yams contributed to the adoption of manioc in the lower Niger because of widespread hunger. Although the Portuguese had introduced manioc to the Niger Delta centuries earlier (Hall 1991), Africans did not choose to use it until the twentieth century because they preferred to cultivate yams (Ohadike 1981). Manioc is now well established as a staple crop in the wetter regions of West Africa (Cock 1985).

Exactly when Europeans imported manioc into East Africa is less certain. It seems improbable that the Portuguese delayed introducing manioc to their colony of Mozambique until 1750, as the historian Justus Strandes claims (Jones 1959). But if Strandes is correct, then the French may have been first to do it when, around 1736, the French governor, Mahé de la Bourdonnais, had manioc brought from Brazil to the island of Mauritius in the Indian Ocean (Pynaert 1951). Two years later, the French Compagnie des Indes introduced manioc to Réunion, an island near the large island of Madagascar across from Mozambique.

The French also sent manioc plants from Brazil to the French islands near Madagascar, where an initial attempt to plant manioc proved a disaster. Africans there were clearly unfamiliar with bitter manioc. When they tried to eat the fruits of their first harvest, they died of poisoning (Hubert and Dupré 1910). After this tragedy, both the French and Africans learned how to process bitter manioc, and they spread the plant and its processing technology to other small islands near Madagascar. At some point, manioc was also transferred from Réunion to Madagascar, where it became a major staple (Fauchère 1914).

During this period while the French were active in the Indian Ocean, the Portuguese, sometime before 1740, may have introduced manioc along with pineapple and corn to the coast of Tanzania and Kenya. Once established on the East African coast, principally the island of Zanzibar and the town of Mozambique, manioc cultivation progressed inland to the great lakes of Tanganyika and Victoria. In the nineteenth century, manioc plants from the east coast met manioc introduced from the Congo Basin. By that time, manioc cultivation had crossed central Africa from west to east, with Africans rather than European colonialists playing the key role in its diffusion (Jones 1959).

The delay in the introduction of manioc to East Africa until the eighteenth century has been attributed by Jones (1959) to the lack of an intensive colonial presence or to a lack of incentive for its introduction because, before 1800, most East African

slaves were exported to Arabia, Persia, or India. Moreover, East Africa lacked dense forests, and much of the terrain was covered with wooded savannah so that environmental conditions were not as propitious as in the Congo basin. Even as late as 1850, Jones (1959: 84) concludes, "manioc was either absent or unimportant in most of East Africa . . . except right along the coast and in the vicinity of Lake Tanganyika." Its cultivation spread thereafter, however, because British and German colonial officers required its planting as a famine reserve – a role it has continued to play to the present day.

Manioc is now cultivated throughout Africa, excluding the desert north and the far south, and its range is still extending, as it is often the only crop that farmers can cultivate under conditions of low soil fertility, drought, and locust plagues. In part, the twentieth-century diffusion of manioc in Africa may be closely connected with recent population growth – but as a consequence, rather than a cause. As more and more people have put pressure on the types of fertile lands suitable for yams, millet, and sorghum, the land-poor have had to turn to manioc to ward off hunger. Thus, if population growth continues, manioc may become even more central than it now is to African economies and diets.

Asia

In addition to diffusing manioc to Africa, Europeans also transported the American crop to Asia, although Polynesians may have also introduced it into the Pacific via Easter Island. The first Asian region to receive manioc from the Europeans was the Philippines (Philippines, 1939). Various terms – *balingboy*, *Kamoteng-Kahoy*, or *Kamoteng-moro*, manioc was brought to the Philippines by early Spanish settlers. Apparently, manioc plants traveled via Spanish ships across the Pacific from Mexico. As in West Africa, however, its cultivation grew slowly, and it was noted in the late 1930s that manioc in the Philippines "has not been as extensive as in other tropical countries" (Philippines 1939: 3). It had, however, evolved as a major crop in the Mindanao area of the Philippines by the 1980s (Centro Internacional de Agricultura Tropical [CIAT] 1986).

By the seventeenth century, manioc was found in the Moluccas, and by 1653 was being grown on Ambon, one of the outer islands of Indonesia (Pynaert 1951; CIAT 1986). It is likely that the Portuguese brought it to Goa (India) in the early eighteenth century. Additional plants were imported to India from South America in 1794, and from the West Indies in 1840 (Cock 1985). As of 1740, manioc was being raised on Java, and plants taken from there were introduced to Mauritius (CIAT 1986) shortly after the French brought the first manioc plants from Brazil. Thus, Brazilian plants met varieties evolved in Asia on a small island in the Indian ocean.

Mauritius then served as a distribution point for manioc into Sri Lanka (formerly Ceylon), the island at the tip of India, where the Dutch governor, Willem Jacob van de Graaff, introduced it in 1786. Subsequent importations were recorded in 1821 and 1917, and these, too, were from Mauritius. Since then, manioc has been cultivated by peasants, and it is "consumed mainly by the poorest people" (CIAT 1986: 115).

By 1800, manioc cultivation in tropical Asia stretched from Ceylon to the Philippines. It had not, however, replaced Asia's main staple, rice, although it was becoming "the most important" of the American crops "in terms of volume produced" (CIAT 1986: 171). An upland crop in tropical Asia, manioc has served to supplement inadequate supplies of rice, and it was most widely accepted in the land-scarce regions of Java in Indonesia and Kerala in southern India. As in Africa in the nineteenth century, those who convinced or required the inhabitants to accept this famine reserve were the European colonialists, in this case the Dutch in Java and the British in India (CIAT 1986).

European colonialists were especially active in diffusing manioc cultivation and processing techniques in the nineteenth century. They first established a processing and export industry in Malaya in the 1850s and subsequently developed a trade in tapioca with Europe (Cock 1985). In 1886, the Singapore Botanic Gardens introduced new manioc plants to Malaysia (Cock 1985). The Dutch followed by transporting manioc plants from their South American colony of Suriname to Java in 1854. By the early twentieth century, manioc was flourishing (Pynaert 1951), and its cultivation has continued to expand since then.

By the 1980s, at least one-fourth of all manioc grown in Asia was located in Indonesia, with the greatest share on Java (CIAT 1986). Sometime around 1850, manioc was introduced into Thailand, where it has become very popular in the eastern seaboard provinces. Since 1956 it has spread to the northeastern, western, and upper-central provinces of Thailand (CIAT 1986).

The British played a major role in the diffusion of manioc cultivation in southern India, where it was most widely accepted, especially in Kerala. Apparently it was introduced quite late to Calcutta (1794) and Serampur in 1840 (Pynaert 1951). Since then, manioc has evolved in India as a supplementary food staple. It is often consumed as a traditional steam-cooked breakfast dish called *puttu*, or marketed for *sago* (tapioca pearl), starch, and cattle feed (CIAT 1986). It is also a staple food in parts of Myanmar (formerly Burma) (Cock 1985).

Five hundred years after the Europeans discovered manioc, the yield of this American crop abroad has surpassed yields in its homeland. In 1982, the world's manioc production was estimated at 129 million tons, with Asia and Africa accounting for about three-fourths of this production and Latin America contributing only one-fourth.

In tropical Asia, manioc production was nearly 46 million tons in 1982, half of which came from Thailand. The two other major Asian producers were Indonesia and India (CIAT 1986). All this followed upon two decades of rapid growth in manioc production with a doubling of output during the period. The most rapid increases occurred in Thailand, in part due to exports to the European Economic Community. Although manioc is also grown in southern China, mainly on the dryland slopes of Guangdong province and the Guangxi Zhuang autonomous region, accurate statistics for the People's Republic of China are hard to come by. The primary use of manioc in China, however, appears to be as animal feed, especially for hogs (CIAT 1986). It is also grown in Taiwan (Cock 1985).

The Pacific Islands

The final region of manioc production, at present, is the Pacific Islands. In general, scholars have maintained that manioc was not introduced to the Pacific Islands until the mid-nineteenth century. Robert Langdon, however, suggests a different history of diffusion to Oceania. He has recently reported the discovery of a Spanish manuscript recording the presence of *yuca* (a word used in Peru for manioc) on Easter Island in 1770. Captained by Felipe Gonzalez, the expedition sailed from Peru, to reach Easter Island in late 1770. Those who went ashore observed fields of *yuca* under cultivation as well as sweet potatoes, another American crop.

Obviously, the question that this raises is how manioc had reached an island 2,000 miles from South America, unless it had been carried there by Amerindians before Columbus, as Thor Heyerdahl has long argued. Certainly Langdon (1988: 324) observes that the presence of manioc on Easter Island in 1770 "greatly strengthens the case for prehistoric American Indian influence on Easter Island and other islands of eastern Polynesia." In any case, it seems that manioc has now been documented in the Pacific 80 years before Captain Louis A. Bonard took the plant to Tahiti in 1850, from which it spread rapidly to other Pacific islands, where it is now cultivated (CIAT 1986; Langdon 1988).

Conclusion

From a present-day perspective, the contribution of the American crop, manioc, to the world's food supply has largely been unheralded except by Brazilians, by a few historians such as William Jones in his classic *Manioc in Africa*, and by French officials such as Paul Hubert and Emile Dupré in *Le Manioc*, which provides a global view of manioc cultivation as of 1910. Historians have recognized the historical significance of other American crops that played a major role in European history, but it may well be

that manioc's historical impact and diffusion have been slighted because it is a tropical, Third World crop.

Nevertheless, manioc has been as significant to the historical evolution of tropical countries, such as Brazil and Zaire, as the potato has been to that of European countries, such as Ireland and Germany. As world population growth continues into the twenty-first century, manioc may assume an even greater role, enabling the rural poor in developing countries to survive hunger and famine. This versatile food crop, which can overcome drought, survive typhoons and locust plagues, and reproduce on marginal soils, may well make a significant difference in population survival in the tropics in the twenty-first century.

Mary Karasch

Bibliography

- Aguiar, Pinto de. 1982. *Mandioca - Pão do Brasil*. Rio de Janeiro.
- Albuquerque, Milton de. 1969. *A mandioca na Amazônia*. Belém, Brazil.
- Brandão Sobrinho, Julio. 1916. *Mandioca*. São Paulo.
- Cascudo, Luis da Camara. 1984. *Dicionário do folclore Brasileiro*. Fifth edition. Belo Horizonte.
- Centro Internacional de Agricultura Tropical (CIAT) and ESCAP Regional Co-ordination Centre for Research and Development of Coarse Grains, Pulses, Roots and Tuber Crops in the Humid Tropics of Asia and the Pacific. 1986. *Cassava in Asia, its potential and research development needs*. Proceedings of a regional workshop held in Bangkok, Thailand, 5-8 June 1984. Cali, Colombia.
- Cock, James H. 1985. *Cassava: New potential for a neglected crop*. Boulder, Colo., and London.
- Coe, Sophie D. 1994. *America's first cuisines*. Austin, Tex.
- Conceição, Antonio José da. 1979. *A mandioca*. Cruz das Almas, Bahia, Brazil.
- Doku, E. V. 1969. *Cassava in Ghana*. Accra, Ghana.
- Fauchère, A. 1914. *La culture du manioc à Madagascar*. Paris.
- Hall, Robert L. 1991. Savoring Africa in the New World. In *Seeds of change*, ed. Herman J. Viola and Carolyn Margolis, 161-71. Washington, D.C., and London.
- Hubert, Paul, and Émile Dupré. 1910. *Le manioc*. Paris.
- Humboldt, Alexander von. [1811] 1988. *Political essay on the kingdom of New Spain*, trans. John Black, ed. Mary Maples Dunn. Norman, Okla., and London.
- Jennings, D. L. 1976. Cassava. In *Evolution of Crop Plants*, ed. N. W. Simmonds, 81-4. London and New York.
- Johns, Timothy. 1990. *With bitter herbs they shall eat it: Chemical ecology and the origins of human diet and medicine*. Tucson, Ariz.
- Jones, William O. 1959. *Manioc in Africa*. Stanford, Calif.
- Karasch, Mary. 1986. Suppliers, sellers, servants, and slaves. In *Cities and society in colonial Latin America*, ed. Louisa Schell Hoberman and Susan Migden Socolow, 251-83. Albuquerque, N.Mex.
- Langdon, Robert. 1988. Manioc, a long concealed key to the

- enigma of Easter Island. *The Geographical Journal* 154: 324–36.
- Lathrap, Donald W. 1970. *The upper Amazon*. New York and Washington, D.C.
- Léry, Jean de. [1578] 1990. *History of a voyage to the land of Brazil, otherwise called America*, trans. Janet Whatley. Berkeley and Los Angeles.
- Levin, Suzanne. 1983. Food production and population size in the Lesser Antilles. *Human Ecology* 11: 321–38.
- Morley, Sylvanus G., and George W. Brainerd. 1983. *The ancient Maya*. Fourth edition, rev. Robert J. Sharer. Stanford, Calif.
- New York World's Fair. 1939. *Manioc*. Brazil: Official Publication. New York.
- Ohadike, D. C. 1981. The influenza pandemic of 1918–19 and the spread of cassava cultivation on the lower Niger: A study in historical linkages. *Journal of African History* 22: 379–91.
- Onwueme, I. C. 1978. *The tropical tuber crops: Yams, cassava, sweet potato, and cocoyams*. Chichester and New York.
- Peckolt, Theodoro. 1878. *Monographia do milbo e da mandioca: Sua historia, variedades, cultura, uso, composição chimica, etc. . . .*, Vol. 3 of *Historia das plantas alimentares e de gozo do Brasil. . .* Rio de Janeiro.
- Philippines, Commonwealth of the. 1939. Department of Agriculture and Commerce. *The cassava industry in the Philippines*. Manila.
- Price, Richard. 1991. Subsistence on the plantation periphery: Crops, cooking, and labour among eighteenth-century Suriname maroons. In *The slaves' economy: Independent production by slaves in the Americas*, ed. Ira Berlin and Philip D. Morgan. Special Issue of *Slavery and Abolition: A Journal of Comparative Studies* 12: 107–27.
- Pynaert, L. 1951. *Le manioc*. Second edition. Brussels.
- Reff, Daniel T. 1998. The Jesuit mission frontier in comparative perspective: The reductions of the Rio de La Plata region and the missions of northwestern Mexico 1588–1700. In *Contested grounds*, ed. Donna J. Guy and Thomas E. Sheridan. 16–31. Tucson, Ariz.
- Relação das guerras feitas aos Palmares de Pernambuco no tempo do Governador D. Pedro de Almeida de 1675 a 1678 [and attached documents]. 1988. In *Alguns documentos para a história da Escravidão*, ed. Leonardo Dantas Silva, 27–69. Recife, Brazil.
- Rocha Pita, Sebastião da. 1976. *História da América Portuguesa*. São Paulo.
- Roosevelt, Anna Curtenius. 1980. *Parmana: Prehistoric maize and manioc subsistence along the Amazon and Orinoco*. New York.
- Staden, Hans. [1557] 1974. *Dois viagens ao Brasil*, trans. Guiomar de Carvalho Franco. Belo Horizonte.
- Tejada, Carlos. 1979. Nutrition and feeding practices of the Maya in Central America. In *Aspects of the history of medicine in Latin America*, ed. John Z. Bowers and Elizabeth F. Purcell, 54–87. New York.
- Tomich, Dale. 1991. Une petite Guinée: Provision ground and plantation in Martinique, 1830–1848. In *The slaves' economy: Independent production by slaves in the Americas*, ed. Ira Berlin and Philip D. Morgan. Special Issue of *Slavery and Abolition: A Journal of Comparative Studies* 12: 68–91.
- Toussaint-Samat, Maguelonne. 1992. *History of food*, trans. Anthea Bell. Cambridge, Mass.
- Wyman, Donald. 1991. Cassava. *The encyclopedia Americana: International edition*, Vol. 5. Danbury, Conn.

II.B.3 🍠 Potatoes (White)

This chapter presents the paradoxical history of the potato (*Solanum tuberosum*) in human food systems. It is now the fourth most important world food crop, surpassed only by wheat, rice, and maize. In five centuries, this diverse and adaptable tuber has spread from its original South American heartland in the high Andes to all elevation zones in temperate regions of all the continents, and, lately, its production has been increasing most rapidly in the warm, humid, tropical Asian lowlands during the dry season (Vander Zaag 1984).

In the course of its history, the potato adapted, and was adopted, as a highland subsistence crop on all continents. In Europe, it was originally an antifamine food but then became a staple. In Africa and Asia, it has been a vegetable or costaple crop. The potato has been credited with fueling the Industrial Revolution in eighteenth-century Europe but blamed for the mid-nineteenth-century Irish famine. Over three centuries, it also became a central and distinctive element of European regional, and then national, cuisines. Although “late blight” has continued to plague those dependent on potatoes for sustenance (CIP 1994), the potato’s popularity has nevertheless grown since the end of World War II, particularly in its forms of standardized industrially produced potato fries, chips, and other frozen and processed “convenience” foods. Acceptance of standard fries (with burgers) and packaged chips symbolizes the “globalization of diet,” as McDonald’s, Pepsico, and other transnational food firms move potatoes around the world yet another time in their successful creation and marketing of a universal taste for these products.



White potato

In addition, the 1972 creation of an International Potato Center (CIP) in Lima, Peru, with its regional networks, has greatly accelerated the introduction of improved potato varieties and supporting technologies throughout the developing world.

R. N. Salaman's monumental volume charted *The History and Social Influence of the Potato* (1949) – a book that was edited and reprinted in 1985 by J. G. Hawkes, who updated archaeological and agronomic histories and then subsequently issued his own study (Hawkes 1990). The archaeological evidence for the origins of potato domestication is still fragmentary (for example, Hawkes 1990). However, collections, characterizations, and taxonomies of both wild and cultivated forms (Ochoa 1962; Huaman 1983; Hawkes and Hjerting 1989) continue to progress and are generating conclusions about evolutionary relationships that can now be tested with additional cytoplasmic and molecular data from crossability trials (Grun 1990). Such conclusions can also be tested by complementary ethnohistorical, social historical, and culinary historical data (Coe 1994).

Recent biological and cultural histories are recounted in several volumes by CIP (CIP 1984; Horton and Fano 1985; Horton 1987; Woolfe 1987), which also issues an *Annual Report* and a *Potato Atlas*. Key breeding and agronomic advances are also reported in *The Potato Journal*, *American Potato Journal*, *European Potato Journal*, *Potato Research*, *Proceedings of the Potato Association of America*, and reports by the potato marketing boards of major producing countries. All are contributing to worldwide understanding and utilization of potatoes, which exhibit perhaps the greatest amount of biodiversity of any major food crop (Hawkes and Hjerting 1989: 3), with matching cultural diversity in food and nonfood uses.

The Potato in South America: Origins and Diffusion

Cultivated potatoes all belong to one botanical species, *Solanum tuberosum*, but it includes thousands of varieties that vary by size, shape, color, and other sensory characteristics. The potato originated in the South American Andes, but its heartland of wild genetic diversity reaches from Venezuela, Colombia, Ecuador, Peru, Bolivia, Argentina, and Chile across the Pampa and Chaco regions of Argentina, Uruguay, Paraguay, and southern Brazil and northward into Central America, Mexico, and the southwestern United States. There are more than 200 wild potato species in this wide habitat that extends from high cold mountains and plateaus into warmer valleys and subtropical forests and drier semiarid intermontane basins and coastal valleys.

The greatest diversity in wild potato species occurs in the Lake Titicaca region of Peru and Bolivia, where the potato probably was domesticated between 10,000 and 7,000 years ago. *Solanum*

tuberosum most likely was domesticated from the wild diploid species *S. stenotomum*, which then hybridized with *S. sparsipilum* or other wild species to form the amphidiploid *S. tuberosum* that evolved from the short-day northern Andean subspecies *andigena*, via additional crosses with wild species, into the subspecies *tuberosum*, which had a more southerly, longer-day distribution (Grun 1990; Hawkes 1990). Frost resistance and additional pest and disease resistance were introduced later via hybridizations with additional wild species, which allowed potatoes to be grown at altitudes up to 4,500 meters.

Archaeological Evidence

Fossilized remains of possibly cultivated tubers found on a cave floor in Chilca Canyon suggest that the potato was cultivated at least from about 7,000 years ago, although it is not possible to tell whether these were wild, "dump heap," or already garden acquisitions (Ugent 1970). Potato remains (along with those of sweet potato and manioc) from Ancon-Chillon (to the north of Lima) date from 4,500 years ago; northern coastal remains from the site of Casma date from between 4,000 and 3,500 years ago (Ugent et al. 1982). It is surmised that cultivated varieties were being planted on terraces at intermediate altitudes, extending from the river valleys into the high mountains, by the middle initial period between 4,000 and 3,500 years ago. Coastal remains from the monumental preceramic site of El Paraiso (3,800 to 3,500 years ago) suggest a mixed subsistence strategy, including unspecified *Solanum* plants that might be potatoes (Quilter et al. 1991).

Art provides additional testimony for the potato's centrality and for the antiquity of processed potatoes in pre-Columbian Andean culture. Fresh and freeze-dried potatoes are depicted in ceramics of the Moche people of northern Peru (A.D. 1 to 600), on urns in Huari or Pacheco styles from the Nazca Valley (650 to 700), and later Chimu-Inca pots (Hawkes 1990). Post-contact-period Inca wooden beakers also depict potato plants and tubers.

South American civilizations and states were based on vertically integrated production and consumption systems that included seed crops (especially maize, secondarily quinoa) at lower altitudes, potatoes and other tubers at higher altitudes, and llamas (camelids) to transport goods between zones. Hillside terracing conserved moisture and soils and encouraged the selection of multiple cultivars of a number of species that fit into closely spaced ecological niches. Ridged, raised, or mounded fields (still used for potato cultivation around Lake Titicaca) were a type of specialized field system that saved moisture and also protected against frost. In addition to making use of short-term storage in the ground, Andean peoples stored potatoes in fresh or processed forms. Huanaco Viejo and other Inca sites reveal extensive tuber storage areas, constructed in naturally cool zones, where indigenous

farmers (or their rulers) stored whole tubers with carefully managed temperature, moisture, and diffused light to reduce spoilage (Morris 1981). Traditional freeze-drying techniques took advantage of night frosts, sunny days, and running water at high elevation zones and allowed potatoes to provide nourishment over long distances and multiple years, as dehydrated potatoes moved from higher to lower altitudes, where they were traded for grain and cloth.

Biocultural Evolution

As South American cultivators expanded into many closely spaced microenvironmental niches, they selected for thousands of culturally recognized potato varieties of differing sizes, colors, shapes, and textures, with characteristics that provided adequate resistance to pests, frost, and other stressors. At higher altitudes, cultivators selected for bitter varieties of high alkaloid content that were detoxified and rendered edible by freeze-drying (Johns 1990). Culturally directed genetic diversification continues up to the present, as Andean farmers allow wild specimens to grow and hybridize with cultivars, conserving biodiversity while diffusing risk (Ugent 1970; Brush 1992).

The botanical history of the cultivated potato is slowly being assembled by considering together the findings from plant scientists' genetic and taxonomic studies, archaeologists' interpretations of archaeological and paleobotanical remains, and ethnographers' observations and analogies from contemporary farming, food processing, and storage. Plant scientists continue to explore wild and cultivated habitats in the potato's heartland, where they find wild potato species that offer a tantalizing range of useful characteristics to protect against frost; against fungal, viral, and bacterial infections; and against nematodes and insects (for example, Ochoa 1962; Ochoa and Schmiediche 1983). Carnivorous, sticky-haired species, such as *Solanum berthaultii*, devour their prey; others repel them pheromonically by mimicking the scent of insects under stress (Hawkes and Hjerting 1989).

Added into the botanical and archaeological data mix are culinary historians' insights from agricultural, botanical, lexical, and food texts. Guaman Poma de Ayala, shortly after Spanish penetration, depicted and described plow-hoe potato and maize cultivation in his chronicle of the Incas (1583–1613) (Guaman Poma de Ayala 1936). Dictionaries that record concepts of the sixteenth-century Aymara peoples from Peru describe time intervals in terms of the time it took to cook a potato (Coe 1994)!

Indigenous peoples also developed detailed vocabularies to describe and classify potatoes, as well as myths and rituals to celebrate the tubers' importance. Even after conversion to Catholicism, they continued to use potatoes in their religious festivals; for example, garlands of potatoes are used to decorate the image of the Virgin Mary at the festival of the Immacu-

late Conception in Juli, Peru (Heather Lechtman, personal communication).

Indigenous Potato Products

Indigenous use of potatoes has included the development of processing methods to extend their nutritional availability and portability. In high altitude zones, selected varieties undergo freezing, soaking, and drying into a product called *chuño* that is without unhealthful bitter glycoalkaloids, is light and easily transported, and can be stored for several years. To render *chuño* (freeze-dried potato), tubers are frozen at night, then warmed in the sun (but shielded from direct rays). Next, they are trampled to slough off skins and to squeeze out any residual water, and then they are soaked in cold running water.

After soaking for 1 to 3 weeks, the product is removed to fields and sun-dried for 5 to 10 days, depending on the cloud cover and type of potato. As these tubers dry, they form a white crust, for which the product is labelled "white *chuño*" (in contrast to "black *chuño*," which eliminates the soaking step). Another processing method involves soaking the tubers without prior freezing for up to a month, then boiling them in this advanced stage of decay. R. Werge (1979) has commented that the odor of this ripening process is "distinctive and strong" and has noted that, as a rule, this product is consumed where it is produced.

Chuño has a long history of provisioning both highland and lowland Andean populations; it was described by early Spanish chroniclers (for example, José de Acosta 1590) and also mentioned in accounts of sixteenth-century mine rations, in which Spanish mine managers complained about its high price. It is curious that one seventeenth-century source mentioned *chuño* as a source of fine white flour for cakes and other delicacies, although it was usually considered to be a lower-class native food (Cobo 1653). Ordinarily, *chuño* is rehydrated in soups and stews.

Another native product is *papa seca* ("dehydrated potato"), for which tubers are boiled, peeled, cut into chunks, sun-dried, and then ground into a starchy staple that is eaten with pork, tomatoes, and onions. *Papa seca* is consumed more widely than *chuño* in urban and coastal areas and can now be purchased in supermarkets.

In areas of frost, potatoes traditionally were also rendered into starch. Traditional products, however, are in decline, as household labor to produce them is now redirected toward higher-value cash employment or schooling. In addition, such traditional products tend to be thought of as inferior, "poor peasant" foods, so that those with cash income and access to store-bought pasta or rice consume these starches instead.

Biodiversity

Declining potato diversity, a byproduct of the insertion of higher-yielding "improved" varieties into South American field systems, is another reason for

the fading of traditional potatoes and potato products. Traditional Andean potato farmers sow together in a single hole as many as 5 small tubers from different varieties and even species, and keep up to 2 dozen named varieties from 3 or 4 species (Quiros et al. 1990; Brush 1992). A particular concern has been whether genetic diversity erodes with the introduction of modern varieties and greater integration of local farmers into regional and national markets. Traditional varieties adapted to lower altitudes (where 75 percent of modern varieties are planted) are at greater risk than those of more mountainous terrains, which are less suited to the cultivation of irrigated, marketable, new varieties. So far, ethnographic investigations do not confirm the conventional wisdom that modern varieties generally compete successfully and eliminate traditional races. Although changes in cropping strategies allocate more land to new, improved varieties, thus reducing the amount of land allocated to traditional varieties, the midaltitude regions that grow modern varieties intensively tend also to devote small areas to older varieties that farmers maintain to meet ritual, symbolic, or preferential local food needs (Rhoades 1984; Brush 1992). In these commercial production zones, the land area allocated to traditional varieties appears to vary with income, with better-off households more likely to maintain larger plots.

Greater production of certain native varieties is actually encouraged by market opportunities. On-farm conservation of potato biodiversity has therefore been favored by the economics of particular native as well as introduced potato varieties, by vertical biogeography, and by persistent cultural customs calling for multiple traditional varieties (Brush, Taylor, and Bellon 1992), and there remains a large amount of as-yet unexploited population variability encoded in folk taxonomies (Quiros et al. 1990). Uniform sowings of improved varieties tend to replace older varieties only in the best-irrigated, midaltitude areas, where farmers harvest and sell an early crop and thus enjoy higher returns for the "new" potatoes. Traditional varietal mixes, however, continue to be grown in higher elevation zones where more extreme and risky environments encourage farmers to propagate a larger variety of them. But unless on-farm conservation programs are encouraged, it may only be a matter of time before erosion occurs.

Andean farmers' ethnotaxonomies ("folk classifications") continue to be studied by anthropologists and plant scientists to learn more about the ways in which traditional peoples recognize and organize plant information. These folk classifications, in most instances, recognize more distinctions than those captured by modern botanical taxonomies, and they also indicate the high value traditional peoples put on maintaining crop species biodiversity as a strategy to reduce risk of total crop failures. The more plant scientists improve their ability to understand the molec-

ular biology, cytology, biochemistry, and genetics of the potato, the more they return to this traditional, natural, and cultural heartland to collect ancient wild and cultivated types and cultural knowledge about how to use potatoes.

In addition, traditional peoples developed ways to store and process potatoes, so that their availability could be extended greatly in time and over space. Agricultural and food scientists, in studying archaeological evidence of cold storage bins (Morris 1981) and contemporary practices (Rhoades 1984), have adopted and disseminated techniques, such as diffused lighting for storage areas and freeze-drying, as ways to increase the potato's food value in other parts of the world. This return to indigenous knowledge at a time of international diffusion of modern molecular technologies is one paradoxical dimension of the potato's history.

The Potato in Europe

Sixteenth-century Spanish explorers, who first observed the potato in Peru, Bolivia, Colombia, and Ecuador, compared the unfamiliar tuber food crop to truffles and adopted the Quechua name, *papa*. The first specimens, arguably short-day *S. tuberosum* ssp. *andigena* forms from Colombia, probably reached Spain around 1570. From there, the potato spread via herbalists and farmers to Italy, the Low Countries, and England, and there was likely a second introduction sometime in the following twenty years. Sir Francis Drake, on his round-the-world voyage (1577 to 1580), recorded an encounter with potatoes off the Chilean coast in 1578, for which British and Irish folklore credits him with having introduced the potato to Great Britain. But this could not have been the case because the tubers would not have survived the additional two years at sea. All European potato varieties in the first 250 years were derived from the original introductions, which constituted a very narrow gene pool that left almost all potatoes vulnerable to devastating viruses and fungal blights by the mid-nineteenth century. *S. tuberosum* ssp. *tuberosum* varieties, introduced from Chile into Europe and North America in the 1800s, represented an ill-fated attempt to widen disease resistance and may actually have introduced the fungus *Phytophthora infestans*, or heightened vulnerability to it. This was the microbe underlying the notorious nineteenth-century Irish crop failures and famine.

Herbal Sources

The potato's initial spread across Europe seems to have involved a combination of Renaissance scientific curiosity and lingering medieval medical superstition. Charles de l'Ecluse or Clusius of Antwerp, who received two tubers and a fruit in 1588 from Philippe de Sivry of Belgium, is credited with introducing the plant to fellow gardeners in Germany,

Austria, France, and the Low Countries (Arber 1938). The Swiss botanist Caspar Bauhin first described the potato in his *Phytopinax* (1596) and named it *Solanum tuberosum* esculentum. He correctly assigned the potato to the nightshade family (Solanum) but otherwise provided a highly stylized, rather than scientific, drawing (1598) and gossiped that potatoes caused wind and leprosy (probably because they looked like leprosy organs) and “incited Venus” (that is, aroused sexual desire), a characterization that led to folkloric names such as “Eve’s apple” or “earth’s testicles.” Such unhealthful or undesirable characteristics probably contributed to potatoes being avoided in Burgundy (reported in John Gerard’s *The Herball*, 1597) and in other parts of Europe. As a result of such persistent negative folklore, the introduction of the potato, a crop recognized by European leaders to have productive and nutritive capacities superior to those of cereal grains (particularly in cold and dry regions), was stymied for years in Germany and Russia.

Gerard – whose printed illustration in his *Herball* of 1597 provided the first lifelike picture of the potato plant, depicting leaves, flowers, and tubers (the plate was revised with careful observation in the later edition of 1633) – appears to have been fascinated by the plant, even wearing a potato flower as his boutonniere in the book’s frontispiece illustration. But he also obscured the true origins of *Solanum tuberosum* by claiming to have received the tubers from “Virginia, otherwise called Norembega,” and therefore naming them “potatoes of Virginia.” The inaccurate name served to distinguish this potato from the “common potato,” *Batata hispanorum* (“Spanish potato”) or *Ipomoea batatas* (“sweet potato”). Additionally, “Virginia” at the time served the English as a generic label for plants of New World (as opposed to European) origin. *The Oxford English Dictionary* contains an entry labeling maize as “Virginia wheat,” although it makes no reference to Gerard’s “potato from Virginia.”

Alternatively, Gerard may have confused a tuber truly indigenous to Virginia, *Glycine apios* or *Apios tuberosa*, with the *Solanum* potato after sowing both tubers together and then attributing an English origin to the tuber of greater significance in order to please his sovereign, Queen Elizabeth (Salaman 1985; Coe 1994). In any case, the false designation and folklore persisted into the next century, by which time potatoes had entered the agricultural economy of Ireland. A legend of Ireland credits the potato’s introduction to the wreck of the Spanish Armada (1588), which washed some tubers ashore (Davidson 1992). Whatever its origins, William Salmon, in his herbal of 1710, distinguished this “Irish” (or “English”) potato from the sweet potato, and “Irish potato” became the name by which “white” (as opposed to “sweet”) potatoes were known in British colonies.

Eighteenth- and Nineteenth-Century Diffusions

The original short-day, late-yielding varieties illustrated in Gerard’s and other herbals had by the eighteenth century been replaced by farmers’ selections for early-maturing varieties that were better suited to the summer day length and climate of the British Isles. The new varieties’ superior yield of calories per unit of land made subsistence possible for small farmers who had lost land and glean rights with the rise of scientific agriculture and the practice of enclosure. Potatoes also provided a new, cheap food source for industrial workers; Salaman (1949), William McNeill (1974), and Henry Hobhouse (1986) were among the historians who saw the potato as having encouraged the rapid rise of population that brought with it the Industrial Revolution.

Potatoes also spread across Italy and Spain. The Hospital de la Sangre in Seville recorded purchases of potatoes among its provisions as early as 1573 (Hawkes and Francisco-Ortega 1992). By 1650, potatoes were a field crop in Flanders, and they had spread northward to Zeeland by 1697, to Utrecht by 1731, to Overijssel by 1746, and to Friesland by 1765. In some high-altitude areas, they were originally adopted as an antifamine food, but the harsh winter of 1740, which caused damage to other crops, hastened potato planting everywhere. By 1794, the tubers had been accepted as an element of the Dutch national dish, a hot pot of root vegetables (Davidson 1992). Toward the end of the eighteenth century, potatoes had become a field crop in Germany, which saw especially large quantities produced after famine years, such as those from 1770 to 1772 and again in 1816 and 1817. Their popularity was increased not only by natural disasters (especially prolonged periods of cold weather) but also by the disasters of wars, because the tubers could be kept in the ground, where stores were less subject to looting and burning by marauding armies.

Such advantages were not lost on such European leaders as Frederick the Great, who, in 1774, commanded that potatoes be grown as a hedge against famine. Very soon afterward, however, potatoes proved to be not so safe in time of war. The War of the Bavarian Succession (1778 to 1779), nicknamed the *Kartoffelkrieg* (“potato war”), found soldiers living off the land, digging potatoes from the fields as they ravaged the countryside. The war ceased once the tuber supply had been exhausted (Nef 1950).

This war in Germany unintentionally provided the catalyst for popularization of the potato in France. A French pharmacist, A. A. Parmentier, had been a German prisoner of war and forced to subsist on potatoes. He survived and returned to Paris, where he championed the tuber as an antifamine food. His promotional campaign saw Marie Antoinette with potato flowers in her hair and King Louis XVI wearing them as boutonnieres. But widespread potato consumption in France still had to wait another century because, at

a time when bread and soup were the French dietary staples, potato starch added to wheat flour produced an unacceptably soggy bread that was too moist to sop up the soup (Wheaton 1983). Widespread utilization of the whole potato in soup or as fries did not occur until well into the following century; even at the time of Jean François Millet's famous "Potato Planters" painting (1861), many French people still considered potatoes unfit for humans or even animals to eat (Murphy 1984).

From the middle eighteenth through nineteenth centuries, potatoes finally spread across central and eastern Europe into Russia. At the end of the seventeenth century, Tsar Peter the Great had sent a sack of potatoes home, where their production and consumption were promoted first by the Free Economic Society and, a century later, by government land grants. But "Old Believers" continued to reject potatoes as "Devil's apples" or "forbidden fruit of Eden," so that as late as 1840, potatoes were still resisted. When, in that year, the government ordered peasants to grow potatoes on common land, they responded with "potato riots" that continued through 1843, when the coercive policy ceased. But, in the next half-century, the potato's obvious superiority to most grain crops and other tubers encouraged its wider growth, first as a garden vegetable and then, as it became a dietary staple, as a field crop (Toomre 1992).

The Social Influence of the Potato

European writers credited the potato with the virtual elimination of famine by the early nineteenth century, without necessarily giving the credit to state political and economic organization and distribution systems (Crossgrove et al. 1990; Coe 1994). Larger-scale potato production subsequently provided surpluses that supported a rise of population in both rural agricultural and urban industrial areas. Potatoes were adopted widely because they grew well in most climates, altitudes, and soils and were more highly productive than grains in both good years and bad. During the seventeenth and eighteenth centuries, selection for earliness and yield gave rise to clones that were better adapted to European temperate, longer-summer-day growing conditions and could be harvested earlier. By the end of the eighteenth century, many varieties were in existence, some specified for human consumption, others as food for animals (Jellis and Richardson 1987). Agricultural workers across Europe increasingly grew potatoes on small allotments to provide food that was cheaper than wheat bread and also inexpensive fodder in the form of substandard tubers. Grains and potatoes, together with the flesh and other products of a few farm animals, provided an economically feasible and nutritionally adequate diet.

No less an authority than Adam Smith, in *An Inquiry into the Nature and Causes of the Wealth of Nations* (1776), estimated that agricultural land allo-

cated to potatoes yielded three times the food/nutrient value of land planted with wheat, so that more people could be maintained on a given quantity of land. Even after workers were fed and the stock replaced, more surplus was left for the landlord. Favorably contrasting the nourishment and healthfulness of potatoes with that of wheat, Smith noted:

The chairmen, porters, and coalheavers in London, and those unfortunate women who live by prostitution, the strongest men and the most beautiful women perhaps in the British dominions, are said to be, the greatest part of them, from the lowest rank of people in Ireland, who are generally fed with the root.

The single outstanding disadvantage of the potato was that stocks could not be stored or carried over from year to year because the tubers rotted (Smith 1776, Volume 1, Book 1, Chapter 11, Part 1: 161-2).

By this time, potatoes were also providing cheap food for growing industrial populations. Low-cost provisions enabled industrialists to keep wages low (Salaman 1985). Indeed, in both rural and urban areas, more than three centuries of resistance to potatoes was overcome. The tuber had been variously regarded as poisonous, tasteless, hard to digest, and an aphrodisiac; some thought of it as pig food, others as famine food or food for the poor, but such prejudices gradually faded as potatoes became the most affordable food staple. Yet, at the same time, the growth of a potato-dependent population in Ireland elicited dire predictions of calamity (by Thomas Malthus, for one), for potatoes were already proving vulnerable to various diseases. Dependent populations were especially at risk because potatoes could neither be stored for more than a few months nor be easily transported into areas of famine, and because those within such populations tended to be politically powerless and economically exploited. For all these reasons, although Ireland suffered a devastating blight that ruined the potato crop from 1845 to 1848, it might accurately be said that the Irish famine was a man-made disaster that could have been prevented or mitigated by timely British emergency relief and greater noblesse oblige on the part of better-off Irish countrymen.

The Potato and Ireland

The history of the potato in Ireland has been summarized by C. Woodham-Smith (1962), A. Bourke (1993), and C. Kinealy (1995), among others. Such accounts trace the way in which the potato, along with the "conacre" system of land and labor allocation and the "lazy-bed" system of potato cultivation, came to dominate Irish agriculture as British landlords made less and less land and time available for their Irish workers' self-provisioning. The advent of more scientifically based agriculture and the enclosure of common lands had left many landless by the end of the eighteenth

century. The “conacre” custom (or economy) allowed landless peasants to rent small plots for 11-month periods in return for agricultural services to the landlord. Peasants managed to feed themselves on such minuscule holdings by setting up raised “lazy” beds in which they placed tubers, then covered them with manure, seaweed, and additional soil to protect them from moisture.

Average yields of potatoes were 6 tons per acre, in contrast with less than 1 ton per acre for wheat or oats. In 1845, the potato crop occupied 2 million acres, and a 13.6 million ton harvest was anticipated, of which slightly less than half would have gone to humans. But grains were higher-value crops, and expansion of roads into the hinterlands during the early decades of the nineteenth century meant that grains could be more easily transported than they previously had been. Thus, values for (grain) export agriculture rose and competed more fiercely with subsistence crops for land. Conacres shrank, and many workers migrated seasonally to Scotland for the harvest, thereby reducing consumption at home and earning additional money for food. This was yet another route by which the potato and its associated social institutions “fed” the industrial economy (Vincent 1995).

“Late blight” (*Phytophthora infestans*), having ravaged potato crops in North America, disrupted this highly vulnerable agro-economic and social context in the 1840s. The blight first appeared in late July 1845 in the Low Countries, spreading from there to England and finally to Ireland, where the poor farming population had no alternative foods to fall back on. It is ironic that late blight probably was introduced into Europe via new potato varieties that had been imported from the Western Hemisphere to counter epidemics of fungal “dry rot” and viral “curl” that had plagued previous decades. Although some scientists had observed that copper sulfate (as a dip for seed or an application for foliage) offered plants protection against what later came to be understood as fungal diseases, the science of plant pathology and pesticides was not yet far advanced, and no preventive or ameliorative steps were taken. “Bordeaux mixture,” an antifungal application suitable for grape vines and potatoes, was not tried until the 1880s.

The blight of 1845 savaged 40 (not 100) percent of the crop, but infected tubers were allowed to rot in the fields, where they incubated the spores of the following years’ disasters. In 1846, ideal weather conditions for late blight aided the rapid infection of early tubers, so that barely 10 percent of the crop was salvaged. But in the aftermath of the less-than-total disaster of 1845, the 1846 emergency was largely ignored by the British government, which failed to suspend the Corn Laws and continued both to export Irish grain and to forbid emergency grain imports. Taxes continued to be enforced, evictions soared, and relief measures, which included food-for-work and soup

kitchens, were too few and too late. Bourke (1993), among others, blamed the English as well as the Irish landlords, a well-off greedy few who benefited from the political and economic policies that impoverished the masses.

Sickness accompanied hunger through 1848, with the result that more than a million and a half Irish people either died or emigrated in search of sustenance. Neither the population nor its potato production ever recovered, although to this day, Ireland’s per capita potato consumption (143 kilograms [kg] per year) surpasses that of rival high consumers in Europe (the Portuguese consume 107 kg per year and Spaniards 106 kg) (Lysaght 1994).

The potato also remains an enduring “polysemous symbol,” celebrated in Irish literature and culinary arts. In the writings of James Joyce, the potato serves as talisman, as signifier of heroic continuity, but also as a symbol of deterioration and decadence (Merritt 1990). Joyce’s references to typical Irish national dishes have been collected, with recipes, into a cookbook entitled *The Joyce of Cooking* (Armstrong 1986).

Later European Developments

European descendants of the original *S. tuberosum* ssp. *andigena* clones were virtually wiped out with the arrival of late blight in the mid-nineteenth century. They were replaced by ssp. *tuberosum* varieties that also – like their predecessors – hybridized readily across subspecies. A single clone, named “Chilean Rough Purple Chili,” has accounted for a large proportion of subsequent European and North American potatoes, including the “Early Rose” and “Russet Burbank” varieties, the latter of which was introduced into the United States in 1876. In addition to Russet Burbank, several very old varieties still predominate in the United States and Europe, notably “Bintje,” introduced into the Netherlands in 1910, and “King Edward,” introduced into the United Kingdom in 1902 (Hermsen and Swiezyński 1987). Attempts to broaden the genetic base for breeding accelerated in the 1920s and 1930s, with N. I. Vavilov’s Russian expedition that collected frost- and blight-resistant varieties from South America and, subsequently, with the British Empire (later Commonwealth) Potato Collecting Expedition (Hawkes 1990).

Blight and viruses notwithstanding, the potato played an ever-expanding role in European food economies. Epitomized in Vincent Van Gogh’s “Potato Eaters” of 1885, but more nobly so in Millet’s “Potato Planters” of 1861, potatoes on the European mainland came to symbolize the rugged, honest peasant, wrestling life and livelihood from the soil. In England, eastern Europe, and Russia, potatoes played significant nutritional roles during ordinary times and assumed extraordinary nutritional roles in war years (Salaman 1985). Even today they remain the fallback crop in times of turmoil, as was seen in Russia in the severe

months of 1992, following glasnost and the reorganization of the economy. An article the same year in the *New Scientist* reported that Russian citizens were planting potatoes everywhere, even illegally in the Losinskii Ostrove National Park, and attempting to steal potatoes from farms!

Europeans were directly responsible for the introduction of potatoes into North America, where they were well established by the eighteenth century. In addition, potatoes accompanied colonists to India, to French Indochina (CIP 1984), to China (Anderson 1988), and to New Zealand where, in the nineteenth century, the Maoris adopted them on the model of other tuber crops (Yen 1961/2). Potatoes also entered Africa with Belgian, British, French, and German colonists, who consumed them as a vegetable rather than as a staple starch. The largest recent expansion of potato cultivation has been in former European colonies, where people in the nineteenth century regarded the tuber as a high-value garden crop and prestigious European vegetable but since then (perhaps in conjunction with the end of colonialism) have come to view it as a staple or costaple garnish and snack (Woolfe 1987).

Potatoes in Developing Countries

In Asia and Africa, the potato has filled a number of production and consumption niches, and its history on these continents has been similar to that in Europe. Once again, despite its advantages as an antifamine, high-elevation alternative to grain, with particular virtues during conflicts, the potato was at first resisted by local farmers, who believed it to be poisonous. In the highest elevation zones, such as the Nepalese Himalayas (Fürer-Haimendorf 1964) and the upper reaches of Rwanda (Scott 1988), potatoes took root as a new staple food crop and contributed to subsistence, surplus, and population expansion. The plants were promoted by savvy rulers, who used demonstration, economic incentives, or coercion to overcome farmers' superstitions and resistance (CIP 1984). In Africa, as in Europe, the popularity of the tubers increased in wartime because they could be stored in the ground.

With the 1972 creation of the International Potato Center (CIP) and its mission to increase potato production and consumption in developing countries while protecting biodiversity, the introduction of improved potato varieties has accelerated around the world. CIP's activities, along with the operation of diverse market forces, have resulted in some African and Asian countries rapidly becoming areas of high potato consumption. Prior to its most recent civil conflict, Rwanda in some localities witnessed per capita consumption as high as 153 to 200 kg per year (Scott 1988) – higher than that in any Western European country, including Ireland. If Rwanda can reattain peace, and agronomic and credit constraints on pro-

duction and infrastructural limits on marketing could be removed, production could expand much farther and faster from the "grassroots," as it has in neighboring Tanzania. There, local farmers in recent years have developed the potato as a cash crop – the result of the introduction of several new varieties brought back by migrant laborers from Uganda, the diffusion of other varieties from Kenya, and the comparative advantage of raising potatoes relative to other cash or subsistence crops (Andersson 1996).

The potato offers excellent advantages as a subsistence crop because of its high yields, low input costs, and favorable response to intensive gardening techniques (for example, Nganga 1984). But potato promotions in Africa ominously echo the terms in which eighteenth- and nineteenth-century British observers praised the tuber. Scientists and political economists should be ever vigilant in ensuring that the potato is not again employed as a stopgap measure in contexts of great social inequality and food/nutritional insecurity, where vulnerability to late blight (or any other stressor) might lead to a repetition of the Great (nineteenth-century Irish) Hunger. Techniques of "clean" seed dissemination and mixed cropping strategies that "clean" the soil are designed to help prevent such calamities now and in the future. But all highlight the need to monitor pests and improve breeding materials so that resistant varieties of the potato are easily available to farmers who have become increasingly reliant on it for food and income.

The same cautions hold for Asia, where production and consumption of potatoes is expanding because of the market as well as international agricultural interests. Since the 1970s, the greatest rate of increase has been in the warm, humid, subtropical lowlands of Asia, where potatoes are planted as a dry-season intercrop with rice or wheat (Vander Zaag 1984), providing income and relief from seasonal hunger (Chakrabarti 1986). The surge in potato production has been spurred in some cases by new seeding materials and techniques. In Vietnam in the 1970s and 1980s, the Vietnamese and CIP introduced superior, blight-resistant clones that could be multiplied by tissue culture and micropropagation methods. Some enterprising farming families then took over the labor-intensive rapid multiplication, so that by 1985, three household "cottage industries" were supplying 600,000 cuttings for some 12,000 farmers (CIP 1984). Production in other Asian nations has also accelerated (for example, in Sri Lanka) as a result of government promotions and policies that have banned imports of all (including seed) potatoes since the 1960s (CIP 1984).

In Central America and the Caribbean, financial incentives and media promotion have been used to increase production and consumption of potatoes in places unaccustomed to them, such as the Dominican Republic, where the state offered credit and guaranteed purchase to potato farmers after the country

experienced a rice deficit (CIP 1984). Similarly, during post-hurricane disaster conditions of 1987, Nicaraguans were encouraged to eat more potatoes – these shipped from friendly donors in the Soviet bloc. In South American countries, campaigns are underway to encourage farmers to grow more potatoes for sale as well as for home consumption, as in Bolivia, where economists hope that as part of diversified employment strategies, an increased production and sale of improved potato varieties can have a multiplier effect, reducing unemployment and increasing access to food (Franco and Godoy 1993). But all of these programs highlight the need to reconcile production and income concerns with the protection of biodiversity and reduction of risks.

Maintaining and Utilizing Biodiversity

Modern scientific attempts to broaden the genetic base for potato breeding began with European scientific expeditions in the 1920s and 1930s, including the already-mentioned Russian (Vavilov 1951) and British collections. Today, major gene banks and study collections are maintained at the Potato Introduction Center, Sturgeon Bay, Wisconsin; the Braunschweig-Volkenrode Genetic Resources Center (Joint German-Netherlands Potato Gene Bank); the N. I. Vavilov Institute of Plant Industry in Leningrad; and the International Potato Center (CIP) in Lima. Major potato-producing countries publish annual lists of registered varieties, standardized to report on agronomic characteristics (disease and pest resistances, seasonality, and environmental tolerances) and cooking and processing qualities (industrial-processing suitability for fries, chips, or dehydration; or home-processing aspects, such as requisite cooking times for boiling, baking, roasting, or frying). Additional consumer descriptors include color, texture, flavor, and the extent of any postcooking tendency to blacken or disintegrate.

Acceptably low alkaloid content is the main chemical toxicity concern, especially because glycoalkaloids are often involved in pest-resistance characteristics introduced during plant breeding. In one historical example, U.S. and Canadian agricultural officials were obliged to remove a promising new multiresistant variety (named “Lenape”) from production because a scientist discovered its sickeningly high alkaloid content (Woolfe 1987).

Since the 1960s, new varieties have been protected by plant breeders’ rights and, internationally, by the *Union Pour la Protection des Obtentions Végétales* (UPOV), which uses a standard set of 107 taxonomic characters to describe individual potato cultivars. UPOV is designed to facilitate exchanges among member countries and so accelerate the breeding process. Collection, conservation, documentation, evaluation, exchange, and use of germ plasm are also regulated by descriptor lists produced in

cooperation with the International Bank for Plant Genetic Resources (IBPGR).

The pace of new varietal introductions is accelerating as more wild species of potential utility for potato improvement are identified and genetically tapped for useful traits that are transferred with the assistance of biotechnology. Wild potatoes with resistance to one pathogen or pest tend to be susceptible to others and may have undesirable growth, tuber, or quality (especially high alkaloid) characteristics. Conventional breeding still requires 12 to 15 years to develop new varieties that include desirable – and exclude undesirable – genes. Protoplast fusion, selection from somaclonal variation, and genetic engineering via *Agrobacterium tumefaciens* are some “unconventional” techniques that promise to widen the scope and quicken the pace of varietal improvement, especially once the genes that control important traits have been characterized. The latter process is facilitated by advances in genetic linkage mapping (Tanksley, Ganal, and Prince 1992) and in practical communication among conventional breeding and agronomic programs (Thomson 1987) that set objectives.

European countries (such as the Netherlands, which has a highly successful seed-potato export business) have been contributing to the development of varieties with superior tolerance for environmental stressors, especially heat and drought, as potato production grows in subtropical countries of Asia and Africa (Levy 1987). Innovative breeding programs also include social components that respond to economic concerns, such as that growing potatoes for market contributes to women’s household income. A Dutch-sponsored program in Asia built up a potato network of women social scientists, nutritionists, and marketing experts along these lines. CIP, in consultation with professionals from national programs, coordinates research and varietal development as well as collection and characterization of germ plasm (seed material) from wild resources.

The Significance of CIP

The International Potato Center (CIP) which grew out of the Mexican national potato program funded by the Rockefeller Foundation, is part of the Consultative Group on International Agricultural Research. It provides a major resource and impetus for strategic studies that tap the genetic and phenotypic diversity of the potato and accelerate the introduction of useful characteristics into new cultivars. Since 1972, CIP has built and maintained the World Potato Collection of some 13,000 accessions, characterized as 5,000 cultivars and 1,500 wild types. In addition to South American programs, CIP potato campaigns extend from the plains of India, Pakistan, and Bangladesh to the oases of North Africa and the highlands and valleys of Central Africa.

CIP’s major technical activities include an effective

population breeding strategy, "clean" (pest- and disease-free) germ-plasm distribution, virus and viroid detection and elimination, agronomy, integrated pest management, tissue culture and rapid multiplication of seed materials, advancement of true potato seed as an alternative to tubers or microtubers, and improvement of storage practices. In the 1990s, a principal thrust of research has been to generate seed materials resistant to late blight, which has reemerged in a more virulent, sexually reproducing form (Niederhauser 1992; Daly 1996).

Strategies involve breeding for multi-gene ("horizontal") rather than single-gene resistance, development and dissemination of true potato seed (which does not disseminate the fungus), and integrated pest management that relies on cost-effective applications of fungicides (CIP 1994). Training, regional networks, and participatory research with farmers are additional dimensions of CIP programs. Collaborative networks offer courses that allow potato specialists to interact and address common problems. In addition, CIP also pioneered "farmer-back-to-farmer" research, whereby effective techniques developed by farmers in one part of the world are shared with farmers in other geographic areas. For example, as already mentioned, reduction of postharvest losses through diffused-light storage is a technique that CIP researchers learned from Peruvian farmers and then brokered to farmers in Asia and Africa (CIP 1984; Rhoades 1984).

CIP also extends its networking to international food purveyors, such as McDonald's and Pepsico - transnational corporations interested in developing improved, pest-resistant, uniformly shaped, high-solid-content potato varieties to be used in making standardized fries. One goal of such enterprises is to develop local sources of supply of raw potatoes for the firms' international franchises, an accomplishment that would improve potato production and income for developing-country farmers and also reduce transportation costs (Walsh 1990). Although principally engaged in agricultural research and extension, CIP also studies consumption patterns, which can improve the potato's dietary and nutritional contributions while eliminating waste (Woolfe 1987; Bouis and Scott 1996).

Dietary and Nutritional Dimensions

Potatoes - simply boiled, baked, or roasted - are an inexpensive, nutritious, and ordinarily harmless source of carbohydrate calories and good-quality protein, and potato skins are an excellent source of vitamin C. Because a small tuber (100 grams [g]) boiled in its skin provides 16 mg of ascorbic acid - 80 percent of a child's or 50 percent of an adult's daily requirement - the potato is an excellent preventive against scurvy. Potatoes are also a good source of the B vitamins (thiamine, pyridoxine, and niacin) and are rich in potassium, phosphorus, and other trace elements. Nutritive value by weight is low because potatoes are

mostly water, but consumed in sufficient quantity to meet caloric needs, the dry matter (about 20 percent) provides the micronutrients just mentioned, an easily digestible starch, and nitrogen (protein), which is comparable on a dry-weight basis to the protein content of cereals and, on a cooked basis, to that of boiled cereals, such as rice- or grain-based gruels (Woolfe 1987).

Potato protein, like that of legumes, is high in lysine and low in sulfur-containing amino acids, making potatoes a good nutritional staple for adults, especially if consumed with cereals as a protein complement. Prepared in fresh form, however, tubers are too bulky to provide a staple for infants or children without an energy-rich supplement. Food technologists are hopeful that novel processing measures may manage to convert the naturally damp, starchy tuber (which molds easily) into a light, nutritious powder that can be reconstituted as a healthful snack or baby food. They also hope to make use of potato protein concentrate, derived either directly by protein recovery or from single-cell protein grown on potato-processing waste (Woolfe 1987). Both advances would minimize waste as well as deliver new sources of nutrients for humans and animals, rendering potato processing more economical. Containing contaminants in industrial potato processing is still very expensive, but sun-drying, a cottage or village industry in India and other Asian countries, holds promise as an inexpensive way to preserve the potato and smooth out seasonal market gluts.

Preservation looms as a large issue because fungus-infected, improperly stored, and undercooked potatoes are toxic for both humans and livestock. Storage and preparation also can diminish the tuber's sensory and nutritional qualities. Sweetening (enzyme conversion of starch), lipid degradation, and discoloration or blackening are signs of deterioration that reduces palatability and protein value. Storage in direct sunlight raises glycoalkaloid content. Other antinutritional factors, such as proteinase inhibitors and lectins, which play a role in insect resistance in some varieties, are ordinarily destroyed by heat, but undercooking, especially when fuel is limited, can leave potatoes indigestible and even poisonous.

Dietary Roles

Although peeling, boiling, and other handling of potatoes decrease micronutrient values, they remove dirt, roughage, and toxins, as well as render potatoes edible. In their Andean heartland, potatoes have always been consumed fresh (boiled or roasted) or reconstituted in stews from freeze-dried or sun-dried forms. They have been the most important root-crop starchy staple, although other cultivated and wild tubers are consumed along with cereals, both indigenous (maize and quinoa) and nonindigenous (barley and wheat). Despite the importance of the potato, cereals were often preferred. For example, Inca rul-

ing elites, just prior to conquest, were said to have favored maize over potatoes, perhaps because the cereal provided denser carbohydrate-protein-fat calories and also was superior for brewing. For these reasons, the Inca may have moved highland peasant populations to lowland irrigated valley sites, where they produced maize instead of potatoes (Earle et al. 1987; Coe 1994). In South America today, potatoes are consumed as a staple or costaple with noodles, barley, rice, and/or legumes and are not used for the manufacture of alcohol. In Central America and Mexico, they are consumed as a costaple main dish or vegetable, in substitution for beans.

In Europe, potatoes historically were added to stews, much like other root vegetables, or boiled, baked, roasted, or fried with the addition of fat, salt, and spices. Boiled potatoes became the staple for eighteenth- and nineteenth-century Irish adults, who consumed up to 16 pounds per person per day in the absence of oatmeal, bread, milk, or pork. These potatoes were served in forms that included pies and cakes (Armstrong 1986). In eastern Europe and Russia, potatoes were eaten boiled or roasted, or were prepared as a costaple with wheat flour in pasta or pastries. In France, by the nineteenth century, fried potatoes were popular, and potatoes were also consumed in soup. In France, Germany, and northern and eastern Europe, potatoes were used for the manufacture of alcohol, which was drunk as a distinct beverage or was put into fortified wines (Bourke 1993). In Great Britain and North America, there developed "fish and chips" and "meat and potatoes" diets. In both locations, potatoes comprised the major starchy component of meals that usually included meat and additional components of green leafy or yellow vegetables.

In former European colonies of Asia and Africa, potatoes were initially consumed only occasionally, like asparagus or other relatively high-cost vegetables, but increased production made them a staple in certain areas. In central African regions of relatively high production, potatoes are beaten with grains and legumes into a stiff porridge, or boiled or roasted and eaten whole. Alternatively, in many Asian cuisines they provide a small garnish, one of a number of side dishes that go with a main staple, or they serve as a snack food consumed whole or in a flour-based pastry. Woolfe (1987: 207, Figure 6.7) has diagrammed these possible dietary roles and has described a four-part "typology of potato consumption" that ranges from (1) potato as staple or costaple, a main source of food energy eaten almost every day for a total consumption of 60 to 200 kg per year; to (2) potato as a complementary vegetable served one or more times per week; to (3) potato as a luxury or special food consumed with 1 to 12 meals per year; to (4) potato as a nonfood because it is either unknown or tabooed. For each of these culinary ends, cultural consumers recognize and rank multiple varieties of potatoes.

Culinary Classifications

In the United States, potato varieties are sometimes classified, named, and marketed according to their geographical location of production (for example, "Idaho" potatoes for baking). They are also classified by varietal name (for example, Russet Burbank, which comes from Idaho but also from other places and is good for baking) and by color and size (for example, small, red, "White Rose," "Gold Rose," "Yukon Gold," or "Yellow Finn," which are designated tasty and used for boiling or mashing). Varieties are also characterized according to cooking qualities that describe their relative starch and moisture content. High-starch, "floury" potatoes are supposed to be better for baking, frying, and mashing; lower-starch, "waxy" potatoes are better for boiling, roasting, and salads (because they hold their shape); and medium-starch, "all-purpose" potatoes are deemed good for pan-frying, scalloping, and pancakes.

Cookbooks (for example, McGee 1984) suggest that relative starch content and function can be determined by a saltwater test (waxy potatoes float, floury varieties sink) or by observation (oval-shaped, thick-skinned potatoes prove better for baking, whereas large, round, thin-skinned potatoes suit many purposes). Specialized cookbooks devoted entirely to the potato help consumers and home cooks make sense of this great diversity (Marshall 1992; see also O'Neill 1992), offering a wide range of recipes, from simple to elegant, for potato appetizers (crepes, puff pastries, fritters, pies, and tarts); potato ingredients, thickeners, or binders in soups; and potato salads, breads, and main courses. They detail dishes that use potatoes baked, mashed, sauteed, braised, or roasted; as fries and puffs (*pommes soufflées* is folklorically dated to 1837 and King Louis Philippe), and in gratinées (baked with a crust); as well as potato dumplings, gnocchi, pancakes, and even desserts. Potato cookbooks, along with elegant presentations of the tubers in fine restaurants, have helped transform the image of the potato from a fattening and undesirable starch into a desirable and healthful gourmet food item.

Mass production over the years has produced larger but more insipid potatoes that are baked, boiled, and mashed, mixed with fats and spices, fried, or mixed with oil and vinegar in salads. Running counter to this trend, however, has been a demand in the 1990s for "heirloom" (traditional) varieties, which increasingly are protected by patent to ensure greater income for their developers and marketers. In the United States, heirloom varieties are disseminated through fine-food stores, as well as seed catalogues that distribute eyes, cuttings, and mini-tubers for home gardens. There is even a Maine-based "Potato of the Month Club," which markets "old-fashioned" or organically grown varieties (O'Neill 1992) to people unable to grow their own.

Breeders are also scrambling to design new gold or purple varieties that can be sold at a premium. In

1989, Michigan State University breeders completed designing a "perfect" potato ("MICHIGOLD") for Michigan farmers: Distinctive and yellow-fleshed, this variety was tasty, nutritious, high yielding, and disease resistant, and (its breeders joked), it would not grow well outside of Michigan's borders (from the author's interviews with Michigan State University scientists). Also of current importance is a search for exotic potatoes, such as the small, elongated, densely golden-fleshed "La Ratte" or "La Reine," which boasts "a flavor that hints richly of hazelnuts and chestnuts" (Fabricant 1996). These return the modern, North American consumer to what were perhaps the "truffle-like" flavors reported by sixteenth-century Spaniards encountering potatoes for the first time. Such special varieties also may help to counter the trend of ever more industrially processed potato foods that has been underway since the 1940s.

Industrially Processed Potato Foods

Since the end of World War II, processed products have come to dominate 75 percent of the potato market, especially as frozen or snack foods. Seventy percent of Idaho-grown and 80 percent of Washington-grown potatoes are processed, and the proportion is also growing in Europe and Asia (Talburtt 1975). Freeze-dried potatoes received a boost during the war years, when U.S. technologists are reported to have visited South America to explore the ancient art of potato freeze-drying and adapt it for military rations (Werge 1979). Since World War II, the development of the frozen food industry and other food-industry processes and packaging, combined with a surging demand for snack and "fast" (convenience) foods, have contributed to the increasing expansion of industrially processed potato products in civilian markets. By the 1970s, 50 percent of potatoes consumed in the United States were dehydrated, fried, canned, or frozen, with close to 50 percent of this amount in the frozen food category. The glossy reports of mammoth food purveyors, such as Heinz, which controls Ore-Ida, proudly boast new and growing markets for processed potatoes (and their standby, ketchup) in the former Soviet Union and Asia.

The other large growth area for fried potatoes and chips has been in the transnational restaurant chains, where fries (with burgers) symbolize modernization or diet globalization. Unfortunately, the "value added" in calories and cost compounds the nutritional problems of consumers struggling to subsist on marginal food budgets, as well as those of people who are otherwise poorly nourished. For less affluent consumers, consumption of fries and other relatively expensive, fat-laden potato products means significant losses (of 50 percent or more) in the nutrients available in freshly prepared potatoes – a result of the many steps involved in storage, processing, and final preparation. Although processed potato

foods are not "bad" in themselves, the marginal nutritional contexts in which some people choose to eat them means a diversion of critical monetary and food resources from more nutritious and cost-effective food purchases. The health risks associated with high amounts of fat and obesity are additional factors.

Potato: Present and Future

Potato consumption is on the rise in most parts of the world. In 1994, China led other nations by producing 40,039,000 metric tons, followed by the Russian Federation (33,780,000), Poland (23,058,000), the United States (20,835,000), Ukraine (16,102,000), and India (15,000,000) (FAO 1995). Average annual per capita consumption is reported to be highest in certain highland regions of Rwanda (153 kg), Peru (100 to 200 kg), and highland Asia (no figures available) (Woolfe 1987), with the largest rate of increase in lowland Asia.

Expansion of potato production and consumption has resulted from the inherent plasticity of the crop; the international training, technical programs, and technology transfer offered by CIP and European purveyors; the ecological opportunities fostered by the "Green Revolution" in other kinds of farming, especially Asian cereal-based systems; and overarching political-economic transformations in income and trade that have influenced local potato production and consumption, especially via the fast-food industry. The use of potatoes has grown because of the ease with which they can be genetically manipulated and because of their smooth fit into multivarietal or multispecies agronomic systems, not to mention the expanding number of uses for the potato as a food and as a raw industrial material.

Genetic Engineering

The potato already has a well-developed, high-density molecular linkage map that promises to facilitate marker-assisted breeding (Tanksley 1992). Coupled with its ease of transformation by cellular (protoplast fusion) or molecular (*Agrobacterium*-assisted) methods, and useful somaclone variants, the potato is developing into a model food crop for genetic engineering. By 1995, there was a genetically engineered variety, containing bt-toxin as a defense against the potato beetle, in commercial trials (Holmes 1995). Where the potato is intercropped rather than monocropped, it also encourages scientists to rethink the agricultural engineering enterprise as a multicropping system or cycle, within which agronomists must seek to optimize production with more efficient uses of moisture, fertilizer, and antipest applications (Messer 1996). Resurgent – and more virulent – forms of late blight, as well as coevolving virus and beetle pests, are the targets of

integrated pest management that combines new biotechnological tools with more conventional chemical and biological ones.

Potatoes continue to serve as a raw material for starch, alcohol, and livestock fodder (especially in Europe). In addition, they may soon provide a safe and reliable source of genetically engineered pharmaceuticals, such as insulin, or of chemical polymers for plastics and synthetic rubbers. Inserting genes for polymer-making enzymes has been the easy step; regulating production of those enzymes relative to natural processes already in the plant is the next, more difficult, one (Pool 1989). A cartoonist (Pool 1989) captured the irony of saving the family farm – whereby small farmers, on contract, grow raw materials for plastics – by portraying the classic Midwestern “American Gothic” farmer husband and wife standing pitchforks in hand, before a field of plastic bottles!

Potato Philanthropy

With less irony, potatoes have come to serve as a model crop for philanthropy. The Virginia-based Society of St. Andrew, through its potato project, has salvaged more than 200 million pounds of fresh produce, especially potatoes, which has been redirected to feed the hungry. Perhaps the memory of Ireland's potato famine continues to inspire acts of relief and development assistance through such organizations as Irish Concern and Action from Ireland, which, along with Irish political leaders (for example, Robinson 1992), reach out to prevent famine deaths, especially as the people of Ireland mark the 150th anniversary of the Great Hunger.

Concluding Paradoxes

In the foregoing history are at least four paradoxes. The first is the potato's transformation in Europe from an antifamine food crop to a catalyst of famine. Ominously, the principal reliance on this species, which makes possible survival, subsistence, and surplus production in high-elevation zones all over the world, and which yields more calories per unit area than cereals, “caused” the Irish famine of 1845–8 and continues to make other poor rural populations vulnerable to famine.

Paradoxical, too, has been the transformation of this simple, naturally nutritious, inexpensive source of carbohydrate, protein, and vitamins into a relatively expensive processed food and less-healthy carrier of fat in the globalization of french fries and hamburgers.

A third paradox is the enduring or even revitalized importance of Andean source materials for the global proliferation of potatoes. Advances in agronomy and varietal improvement have made the potato an increasingly important and diverse crop for all scales and levels of production and consumption across the globe. But in the face of such geographic and culinary developments, the traditional South

American potato cultures continue to present what to some scientists is a surprising wealth of biological, ecological, storage, and processing knowledge (Werge 1979; Rhoades 1984; Brush 1992). The management of biological diversity, ecology of production, and storage and processing methods are three areas in which indigenous agriculture has continued to inform contemporary potato research. Thus, despite dispersal all over the globe, scientists still return to the potato's heartland to learn how to improve and protect the crop.

A fourth paradox is that potatoes may yet experience their greatest contribution to nutrition and help put an end to hunger, not directly as food but as a component of diversified agro-ecosystems and an industrial cash crop. Since their beginnings, potatoes have always formed a component of diversified agro-livestock food systems. Historically, they achieved their most significant dietary role when grown in rotation with cereals (as in Ireland). Today, they are once again being seasonally rotated within cereal-based cropping systems. Because potatoes are intercropped, they stimulate questions about how biotechnology-assisted agriculture can be implemented more sustainably. So far, plant biotechnologists have considered mainly the host-resistance to individual microbes or insects, and never with more than one crop at a time. But adding potatoes to cereal crop rotations encourages scientists, as it does farmers, to look more closely at the efficiency with which cropping systems use moisture and chemicals, and to ask how subsequent crops can utilize effectively the field residues of previous plantings in order to save water and minimize pollution.

Efforts to integrate potatoes into tropical cropping systems, particularly those in the tropical and subtropical lowlands of southern and southeastern Asia, are stimulating such inquiries. Thus, potatoes, perhaps the first crop cultivated in the Western Hemisphere, are now contributing to a revolution of their own in the newest agricultural revolution: the bio- or gene revolution in Asia. In addition, potatoes may also help to save family farms in the United States and Europe by providing income to those growing the crop for plastic.

Ellen Messer

Bibliography

- Acosta, José de. [1590] 1880. *Historia natural y moral de las Indias*, trans. E. Grimston, London, 1604. Hakluyt Society reprint, 1880. Seville.
- Anderson, E. N. 1988. *The food of China*. New Haven, Conn.
- Andersson, J. A. 1996. Potato cultivation in the Uporoto mountains, Tanzania. *African Affairs* 95: 85–106.
- Arber, A. 1938. *Herbals*. Cambridge.
- Armstrong, A. 1986. *The Joyce of cooking: Food and drink from James Joyce's Dublin*. Barrytown, N.Y.

- Bauhin, C. 1596. *Phytopinax*. Basel.
1598. *Opera Quae Extant Omnia*. Frankfurt.
- Bouis, H. E., and G. Scott. 1996. *Demand for high-value secondary crops in developing countries: The case of potatoes in Bangladesh and Pakistan*. International Food Policy Research Institute, Food and Consumption Division Discussion Paper. Washington, D.C.
- Bourke, A. 1993. "The visitation of God"? *The potato and the great Irish famine*. Dublin.
- Braun, J. von, H. de Haen, and J. Blanken. 1991. *Commercialization of agriculture under population pressure: Effects on production, consumption, and nutrition in Rwanda*. Washington, D.C.
- Brush, S. 1992. Reconsidering the Green Revolution: Diversity and stability in cradle areas of crop domestication. *Human Ecology* 20: 145-67.
- Brush, S., J. E. Taylor, and M. Bellon. 1992. Technology adoption and biological diversity in Andean potato agriculture. *Journal of Development Economics* 39: 365-87.
- Chakrabarti, D. K. 1986. Malnutrition: More should be made of the potato. *World Health Forum* 7: 429-32.
- CIP (International Potato Center). 1984. *Potatoes for the developing world: A collaborative experience*. Lima.
1994. *CIP Annual Report 1993*. Lima.
- Cobo, B. [1653] 1890-1893. *Historia del nuevo mundo*, ed. M. Jimenez de la Espada. 4 vols. Seville.
- Coe, S. 1994. *America's first cuisines*. Austin, Tex.
- Crossgrove, W., D. Egilman, P. Heywood, et al. 1990. Colonialism, international trade, and the nation-state. In *Hunger in history*, ed. L. Newman, 215-40. New York.
- Daly, D. C. 1996. The leaf that launched a thousand ships. *Natural History* 105: 24, 31.
- Davidson, A. 1992. Europeans' wary encounter with tomatoes, potatoes, and other New World foods. In *Chillies to chocolate: Food the Americas gave the world*, ed. Nelson Foster and L. S. Cordell, 1-14. Phoenix, Ariz.
- Drake, F. [1628] 1854. *The world encompassed*, ed. W. S. W. Vaux. London.
- Earle, T., ed. 1987. *Archaeological field research in the Upper Mantaro, Peru, 1982-83. Investigations of Inca expansion and exchange*. Los Angeles.
- Fabricant, F. 1996. French revolution in potatoes comes to America. *New York Times*, 25 September: C6.
- Food and Agriculture Organization of the United Nations (FAO). 1995. *FAO Production*, Vol. 48. Rome.
- Franco, M. de, and R. Godoy. 1993. Potato-led growth: The macroeconomic effects of technological innovations in Bolivian agriculture. *Journal of Development Studies* 29: 561-87.
- Fürer-Haimendorf, C. von. 1964. *The Sherpas of Nepal: Buddhist highlanders*. London.
- Gerard, John. 1633. *The herball on general historie of plantes*. London.
- Grun, P. 1990. The evolution of cultivated potatoes. *Economic Botany* 44 (3rd supplement): 39-55.
- Guaman Poma de Ayala and Felipe Guaman. 1936. Nueva cronica y buen gobierno. In *Travaux et Mémoires de l'Institut d'Ethnologie* 23, ed. P. Rivet. Paris.
- Hawkes, J. G. 1990. *The potato: Evolution, biodiversity, and genetic resources*. Washington, D.C.
- Hawkes, J. G., and J. Francisco-Ortega. 1992. The potato during the late 16th century. *Economic Botany* 46: 86-97.
- Hawkes, J. G., and P. P. Hjerting. 1989. *The potatoes of Bolivia: Their breeding value and evolutionary relationships*. Oxford.
- Hermesen, J. G. Th., and K. M. Swiezynski. 1987. Introduction. In *The production of new potato varieties: Technological advances*, ed. G. J. Jellis and D. E. Richardson, xviii-xx. Cambridge.
- Hobhouse, H. 1986. *Seeds of change: Five plants that transformed the world*. New York.
- Holmes, B. 1995. Chips are down for killer potato. *New Scientist* 146: 9.
- Horton, D. 1987. *Potatoes. production, marketing, and programs for developing countries*. Boulder, Colo.
- Horton, D. E., and H. Fano. 1985. *Potato atlas*. International Potato Center. Lima.
- Huaman, Z. 1983. *The breeding potential of native Andean cultivars*. Proceedings, International Congress: Research for the potato in the year 2000, 10th anniversary, 1972-82. Lima.
- Huaman, Z., J. R. Williams, W. Salhuana, and L. Vincent. 1977. *Descriptors for the cultivated potato*. Rome.
- Jellis, G. J., and D. E. Richardson. 1987. The development of potato varieties in Europe. In *The production of new potato varieties: Technological advances*, ed. G. J. Jellis and D. E. Richardson, 1-9. Cambridge.
- Johns, T. 1990. *With bitter herbs they shall eat it: Chemical ecology and the origins of human diet and medicine*. Tucson, Ariz.
- Kinealy, C. 1995. *The great calamity: The Irish famine 1845-52*. Boulder, Colo.
- Levy, D. 1987. Selection and evaluation of potatoes for improved tolerance of environmental stresses. In *The production of new potato varieties: Technological advances*, ed. G. J. Jellis and D. E. Richardson, 105-7. Cambridge.
- Lysaght, P. 1994. *Aspects of the social and cultural influence of the potato in Ireland*. Paper presented at the 10th Internationale Konferenz für Ethnologische Nahrungsforschung, Kulturprägung durch Nahrung: Die Kartoffel, 6-10 June 1994.
- Marshall, L. 1992. *A passion for potatoes*. New York.
- McGee, H. 1984. *On food and cooking*. New York.
- McNeill, W. H. 1974. *The shape of European history*. New York.
- Merritt, R. 1990. Faith and betrayal, the potato: *Ulysses*. *James Joyce Quarterly* 28: 269-76.
- Messer, E. 1996. Visions of the future: Food, hunger, and nutrition. In *The hunger report: 1996*, ed. E. Messer and P. Uvin, 211-28. Amsterdam.
- Morris, C. 1981. Tecnología y organizacion Inca del almacenamiento de víveres en la sierra. In *La Tecnología en el Mundo Andino*, ed. H. Lechtman and A. M. Soldi, 327-75. Mexico.
- Moscow's forest falls to hungry potato eaters. 1992. *New Scientist* 134 (April 4): 6.
- Murphy, A. 1984. *Millet*. Boston.
- Nef, J. U. 1950. *War and human progress*. Cambridge.
- Nganga, S. 1984. The role of the potato in food production for countries in Africa. In *Potato development and transfer of technology in tropical Africa*, ed. S. Nganga, 63-9. Nairobi.
- Niederhauser, J. S. 1992. The role of the potato in the conquest of hunger and new strategies for international cooperation. *Food Technology* 46: 91-5.
- Ochoa, C. M. 1962. *Los Solanum Tuberiferos silvestres del Peru (Secc. Tuberarium, sub-secc. Hyperbasarthrum)*. Lima.
- Ochoa, C., and P. Schmiediche. 1983. Systemic exploitation and utilization of wild potato germplasm. In *Research for the potato in the year 2000*, ed. W. J. Hooker, 142-4. Lima.
- O'Neill, M. 1989. Potatoes come to power. *New York Times*, 27 September: C1, C10.

1992. Hot potatoes. *New York Times Magazine*, March 29.
- Pool, R. 1989. In search of the plastic potato. *Science* 245: 1187-9.
- Quilter, J., B. Ojeda E., D. Pearsall, et al. 1991. Subsistence economy of El Paraiso, an early Peruvian site. *Science* 251: 277-83.
- Quiros, C. F., S. B. Brush, D. S. Douches, et al. 1990. Biochemical and folk assessment of variability of Andean cultivated potatoes. *Economic Botany* 44: 254-66.
- Rhoades, R. E. 1984. *Breaking new ground: Agricultural anthropology*. Lima.
- Robinson, Mary. 1992. *A voice for Somalia*. Dublin.
- Ross, H. 1979. Wild species and primitive cultivars as ancestors of potato varieties. In *Broadening the genetic base of crops*, ed. A. C. Zeven and A. M. van Harten, 237-45. Wageningen, Netherlands.
- Salaman, R. N. [1949] 1985. *The history and social influence of the potato*, ed. J. G. Hawkes. Cambridge.
- Salmon, W. 1710. *The English herbal*. London.
- Scott, G. J. 1988. *Potatoes in central Africa: A survey of Burundi, Rwanda, and Zaire*. International Potato Center. Lima.
- Smith, Adam. [1776] 1904/1950. *An inquiry into the nature and causes of the wealth of nations*, ed. E. Cannon. London.
- Talbur, William S. 1975. *Potato processing*. Third edition. Westport, Conn.
- Tanksley, S. D., M. W. Ganal, and J. P. Prince. 1992. High density molecular linkage maps of the tomato and potato genomes. *Genetics* 132: 1141-60.
- Thomson, A. J. 1987. A practical breeder's view of the current state of potato breeding and evaluation. In *The production of new potato varieties: Technological advances*, ed. D. J. Jellis and D. E. Richardson, 336-46. Cambridge.
- Toomre, J. 1992. *Classic Russian cooking: Elena Molokhovets's gift to young housewives, 1861-1917*. Bloomington, Ind.
- Ugent, D. 1970. The potato. *Science* 170: 1161-6.
- Ugent, D., Tom Dillehay, and Carlos Ramirez. 1987. Potato remains from a late Pleistocene settlement in south central Chile. *Economic Botany* 41: 17-27.
- Ugent, Donald, Sheila Pozorski, and Thomas Pozorski. 1982. Archaeological potato tuber remains from the Cosma Valley of Peru. *Economic Botany* 36: 182-92.
- Vander Zaag, P. 1984. One potato, two potato. *Far Eastern Economic Review* n. vol. (August 23): 64-6.
- Vavilov, N. I. 1951. The origin, variation, immunity and breeding of cultivated plants, trans. K. Starr Chester. *Chronica Botanica* 13: 1-366.
- Vincent, J. 1995. Conacre: A re-evaluation of Irish custom. In *Articulating hidden histories: Exploring the influence of Eric Wolf*, ed. J. Schneider and R. Rapp, 82-93. Berkeley, Calif.
- Walsh, J. 1990. In Peru, even potato research is high risk. *Science* 247: 1286-7.
- Werge, R. 1979. Potato processing in the central highlands of Peru. *Ecology of Food and Nutrition* 7: 229-34.
- Wheaton, B. 1983. *Savoring the past: The French kitchen and tables from 1300 to 1789*. Philadelphia, Pa.
- Woodham-Smith, C. 1962. *The great hunger: Ireland 1946-1949*. London.
- Woolfe, J. A., with S. V. Poats. 1987. *The potato in the human diet*. New York.
- Yen, D. E. 1961/2. The potato in early New Zealand. *The Potato Journal* (Summer): 2-5.

II.B.4 🌴 Sago

Sago is an edible starch derived from the pith of a variety of sago palms, but mostly from two species of the genus *Metroxylon* - *M. sagu* and *M. rumphii*. The sago palms flower only once (hapaxantic) and are found in tropical lowland swamps. Other genera of palms that yield sago starch include *Arecastrum*, *Arenga*, *Caryota*, *Corypha*, *Eugeissona*, *Mauritia*, and *Roystonea*. In all, there are about 15 species of sago palms distributed in both the Old World and the New, with the most significant of these, *M. sagu*, located mainly on the islands of the Malay Archipelago and New Guinea. As a staple foodstuff, only the *Metroxylon* genus appears to be in regular use, generally among populations located in coastal, lacustrine, or riverine areas. Worldwide, sago provides only about 1.5 percent of the total production of starch and, consequently, is fairly insignificant as a global food source (Flach 1983). It is processed into flour, meal, and pearl sago, and is often used for thickening soups, puddings, and other desserts.



Prickly Sago palm

Sago starch is extracted in a variety of ways, although the general process is similar from area to area. The trunk of a felled palm is chopped into sections and then split vertically to allow the pith to be removed. The extracted pith is ground and then repeatedly washed and strained. The strained material is allowed to dry, and the result is pellets of sago starch. When processed in this manner, the average yield of one palm (of 27 to 50 feet meters in height) generally ranges between 130 and 185 kilograms (kg) of sago, which can feed a family of between two and four persons for up to three months.

History, Cultivation, and Production

History

The early history of sago palm use as a food is still unclear. Ethnologists and anthropologists have generally relied on native myths and legends to judge when it was introduced into the diets of many groups worldwide. Some, such as E. Schlesier and F. Speiser, have tended to believe that the sago palm has been utilized as a food source in the Pacific islands since prehorticultural days. J. B. Avé (1977), for example, has stated that Neolithic and Mesolithic artifacts found in insular Southeast Asia included tools used in sago preparation. Although this suggests that sago has been cultivated since ancient times, paleohistorians are not so sure. E. Haberland and others, for example, have contended that sago consumption was a postagricultural development (Ruddle et al. 1978).

By most accounts, the sago palm was essential to the early inhabitants of Southeast Asia, and was probably one of the first plants they exploited as part of their subsistence strategy (Avé 1977; Rhoads 1982; Flach 1983). Geographer Carl O. Sauer believed that the plant's domestication took place there, where people in freshwater areas were able to employ native palms in a variety of ways, including the production of starch, drugs, and fish poisons, as well as fishing nets and lines (Isaac 1970). According to the folk history of the Melanau of Sarawak, the tribe has "always eaten sago," even though they claim that rice, not sago, is their staple food (Morris 1974).

Sago, however, has also been an important food source for peoples in other parts of the world. Evidence, although limited, indicates that during the Chinese Tang Dynasty (618 to 907), sago starch from palms grown in southeast China came to rival milled grain for use in making cakes. Additionally, the nutritive value of *Metroxylon* sago was discussed in the *Pen Ts'ao Kang mu* (The Great Herbal), and *Caryota* palms are mentioned in Ki Han's *Nan Fang Ts'ao Mu Chuang* ("Account of the Flora of the Southern Regions") (Ruddle et al. 1978). For the peoples of the Southwest Pacific, sago palms have been important from ancient times to the present; stands of *M. sagu* and *M. rumphii* have provided staple foods over the centuries for many millions of people (McCurrach 1960).

In the Western Hemisphere, the use of sago starch has been less common, although *Arecastrum romanzoffianum*, *Mauritia flexuosa*, and *Roystonea oleracea* are all varieties that have provided nutritional relief during times of food scarcity. For example, many Paraguayan peasants are said to have survived on sago in the 1870s, following the devastation wrought by the war of the Triple Alliance. And some peoples, such as the Warao Indians of Venezuela, continue to utilize *M. flexuosa* as a dietary staple (Ruddle et al. 1978).

Properties

Sago palms generally reach maturity at about 15 years of age, at which time the tree develops its enormous mass of pith. The pith makes up approximately 68 to 74 percent of the total weight of the tree, whereas the starch content of the pith constitutes about 25 to 30 percent of the pith weight. Raw sago from *Metroxylon* spp. will yield a range of approximately 285 to 355 calories per 100 grams. Nutritionally, about 70 to 90 percent of raw sago is carbohydrate, 0.3 percent is fiber, and 0.2 percent is protein. Although it has a negligible fat content, sago does contain various minerals, including calcium (10 to 30 milligrams [mg]), phosphorus (approximately 12 mg), and iron (0.7 to 1.5 mg) (Peters 1957; Barrau 1960; Platt 1977; Ruddle et al. 1978).

Sago supplies energy needs, but because it is deficient in most other nutrients, its consumption must be complemented with other foods that yield good-quality proteins as well as a range of vitamins. Climate and other environmental factors generally dictate the supplements. In some areas of New Guinea, for example, the inhabitants use leaves and other greens (sometimes grown on platforms raised above water or in limited garden space) along with the products of fishing and hunting. Another source of animal protein is the sago grub, especially for those groups located inland from the wetlands and rivers. Still others have supplemented their diet with coconuts, tubers, roots, and pulses, in addition to greens (Barrau 1960).

Location

The first Western description of sago consumption appears to be that penned by Marco Polo during his travels to Indonesia in the thirteenth century. Polo wrote "Of the Sixth Kingdom, named Fanfur, where Meal is procured from a certain Tree," with the "meal" in question clearly sago starch. A few centuries later, S. Purchas, during his travels in Sumatra, also mentioned sago as a food source (along with rice and millet) (Ruddle et al. 1978). In 1687, W. Dampier noted that sago was one of the more common foods at Mindanao (Tan 1980).

Toward the end of the nineteenth century, sago palms were observed in a number of regions of the world, and Ceram, Borneo, and Sarawak were mentioned as areas of starch production. Today, a survey of

sago use would encompass a vast area, ranging over Malaysia and the Pacific Islands (Boulger 1889; Flach 1983).

Sago is fairly common in the western Pacific, where cultivated stands of the palm cover an estimated 10,000 hectares (Firth 1950). It is also present throughout much of the southwestern Pacific area. In Papua New Guinea, for example, there are approximately 1,000,000 hectares of wild sago stands and 20,000 hectares of cultivated stands. Similarly, in Indonesia there are 1 million hectares of wild stands and 128,000 hectares that are cultivated. Rounding out the major areas of sago palm stands are Malaysia with 33,000 hectares of cultivated stands and Thailand and the Philippines with 5,000 hectares each (Flach 1983).

Unlike most plants, the sago palm has not been geographically dispersed, and in experimenting with ways to introduce this crop to new areas, M. Flach (1983) discovered a number of possible reasons for the failure of previous attempts. His own efforts failed in Surinam, probably the result of inadequate care of the plants. An attempt in the south Sudan also failed, most likely because of that region's low humidity. Flach did have success in Vietnam, where a sago palm stand succeeded at Can Tho. But, as he discovered, there are two additional factors that make it difficult to disperse sago palms. One is the problem of obtaining generative material, and the other is the cumbersome size of the vegetative material (Flach 1983).

Moreover, depending on location, the peoples of the different sago palm regions of the world call the palms by a great variety of names. The Papuans, for example, have 23 names for *Metroxylon* and sago. In pidgin English, it is *saksak*. In other areas of New Guinea, the sago palm is known as *abia*, *aisai*, *akiri*, *ambe*, *api*, *baiao*, *balega*, *barian*, *da*, *dou*, *fi*, *ipako*, *na*, *nafa*, *ndana*, *no*, *poi*, *pu*, and *wariani*. In the New Hebrides, it is known as *natangora*. In the Fiji Islands, sago is referred to as *ota* or *oat* and as *sogo* or *soqa*, and in the Moluccas it is *lapia*. In Indonesia, sago is known as *rambia*, *rembia*, *rembi*, and *rumbia*, along with other similar cognates (Barrau 1960).

Scientific Description

The most important palm trees in sago production are from the genus *Metroxylon*, a term that comes from the Greek words *metra*, meaning "heart of a tree," and *xylon*, meaning "wood" (Whitmore 1973). *Metroxylon sago* and *Metroxylon rumphii* are economically the most important species in the genus (Flach 1983) and appear to be closely related, as they are found in wild stands mixed together with what appear to be intermediates (Flach 1983).

M. sago and *M. rumphii* share a great number of characteristics, as well as the common name "sago palm." It is thought that *M. rumphii* originated in Malaysia, New Guinea, and Fiji, whereas *M. sago* originated in western New Guinea and the Moluccas. The

trunks of the two species reach a height of approximately 10 to 15 meters and are generally about 45 centimeters (cm) in diameter (McCurrach 1960; Whitmore 1973). Their leaves grow to 600 or more centimeters in length, and the leaflets are about 60 to 120 cm long and 2.5 to 7.6 cm broad. The flower stalk, which appears above the leaves, is 4 to 5 meters in length. The flower stalk of *M. rumphii* is black and covered with spines, whereas that of *M. sago* lacks spines. The fruit produced is spherical, dull yellow, and about 5 cm in diameter. The growth cycle of the sago palm ranges from 8 to 17 years (McCurrach 1960; Flach 1983). Although their ideal temperature range has not yet been determined, it is known that sago palms thrive in areas where the temperature only occasionally drops below 15° C.

What is known about the natural habitat of the sago palms has been gleaned largely from observations in environments where they now grow. Indeed, with so little information available, scientists have been forced to study conditions in the natural habitat as well as the centers of cultivation to glean what they can (Flach 1983). Typical of natural habitats are the swamp forests of sago palms in New Guinea, where fresh water is abundant (Barrau 1960).

Outside of swamps, if a climate is too wet, grasses tend to take over and limit propagation. If, on the other hand, the climate is too dry, taller trees will win in competition with the sago palm. It has been suggested that sago palms might survive under drier conditions if well tended. Although sago palms are relatively tolerant of salinity, if the water becomes too brackish, other trees in the vicinity, such as the nipa palm (*Nipa frutescens*), tend to take over the swamp (Ruddle et al. 1978).

To the conditions of the sago palm's natural habitat, Rhoads (1982) has added the proviso that they are generally "alluvial freshwater swamps" that are frequently located inland from the mouths of large rivers. He has also noted that the mineral soils in which sago palms grow best, especially those high in organic content, need regular flooding for the consistent replacement of nutrients and to discourage the growth of taller trees that keep out sunlight (Rhoads 1982).

Numerous other palms can be sources of sago starch, but they are not so fruitful as *M. sago* and *M. rumphii*. G. S. Boulger (1889) noted that "inferior" sago could generally be obtained from the Gomuti palm (*Arenga saccharifera*), the Kittool palm (*Caryota urens*), and the Cabbage palm (*Corypha umbraculifera*). In the East Indies, sago could be gotten from *Raphia flabelliformis*, *Phoenix farinifera*, and *M. filare*, and in South America from *Mauritia flexuosa* and *Guilielma speciosa* (Boulger 1889). There are also a number of *Metroxylon* species in Oceania, including *amicarum*, *bougainvillense*, *warburgii*, *vitiense*, *upolense*, *salmonense*, and, presumably, *oxybracteatum* (Ohtsuka 1983).

In South America, additional sago-producing palms have been identified among four different genera: *Syagrus*, *Copernicia*, *Mauritia*, and *Manicaria*; moreover, many South American tribes have extracted sago from *Syagrus romanzoffianum* and *Copernicia cerifera* (Wilbert 1976).

Sago Palm Grove Management

Rhoads (1982) determined three general methods of sago palm grove management. The first is simply the process of harvesting sago trees for starch, which (even if only an unintended result) does increase the vitality of the grove: The cutting of palm trunks allows more sunlight to reach nearby shoots, a process that enhances growth and helps to ensure the maturation of at least one sucker, and the damage caused during harvesting by fallen palms and by the construction of work sites in the grove tends to give young sago palm shoots advantages over competitors. Such "unintended management" can be very important to the maintenance and promotion of a sago palm grove (Rhoads 1982).

A second process of sago palm management, termed "horticulture" by Rhoads (1982), involves the planting of suckers or the nurturing and replanting of seedlings. This method, however, is either rare or poorly documented.

A third method of "palm cultivation" involves both the planting of suckers and conscious efforts to change the environment in ways that will promote sago palm growth. One process in which the environment is changed is the creation of artificial swamps by damming streams to flood the groves. Another, observed by Rhoads, is the clearing of the canopy of higher trees to promote sago palm growth. Groves are also sometimes laid out higher up on the slopes of mountains to provide more sunlight for the palms (Rhoads 1982).

Generic Sago Extraction Process

Although sago extraction methods differ somewhat throughout cultures and regions, there are procedures common to all. At the "domestic level," the entire process of sago extraction takes place in the grove itself, thus eliminating the need to transport heavy palm trunks (Flach 1983). Felling the tree occurs when the flowering of the palm indicates that the starch content is at a maximum (Flach 1983). It is also possible to estimate the starch content by taking a small slice from the palm trunk and sampling the starch, either by chewing the pith or by allowing the starch to dry on the axe. If the starch content is too low to merit harvesting the palm, the sample hole is patched with mud (Flach 1983).

If the palm is ready for harvesting, it is felled with an axe, after which the trunk is split lengthwise. (In an alternative method, only the bark is split - and removed.) The pith, when exposed, is "rasped" with a chopper or small hoe (Flach 1983). In the past, choppers were often constructed out of bamboo, but mod-

ern choppers are more generally made of metal. The pith is rasped at a straight angle to the fiber while the worker sits on the trunk. The resulting mixture of fiber and rasped pith is next placed on a kind of trough made from palm leaves that has a sieve attached to the lowest end (Flach 1983).

At this point, water is added and kneaded into the mixture to start it flowing, whereupon fibers are caught by the sieve while the starch, suspended in water, flows through the sieve and is collected in a tank, perhaps an old canoe. The starch eventually settles to the bottom, whereas the extra water flows over the side of the tank. The fibrous materials are given to pigs, ducks, and chickens to consume. With this process, it is possible to produce approximately 1 kg of dry starch per hour (Flach 1983).

Larger, although still "small-scale," extraction operations require waterways to transport the sago palm trunks to a processing plant. There they are cut into sections of about 1 to 1.2 meters in length that are easier to work with than entire trunks (Flach 1983). Extraction methods employed at such facilities follow the general model already outlined, although at times different instruments and processes are employed (Flach 1983).

Rasping, for example, is done with a variety of instruments. A board with nails driven through it is sometimes used, but there are also numerous types of engine-powered rasps. At times, a "broad side rasper," which runs parallel to the bark, is employed (Flach 1983).

The kneading and sieving process also varies at the extraction plants. At some, the mixture is trampled, whereas at others a slowly revolving mesh washer constructed of wood or metal is used. Still other plants employ horizontal screen washers or spiral screw washers. It is also possible to combine mechanical stirring with a mesh washer to process the overflow (Flach 1983).

Small ponds are often utilized for the settling process, although another method involves "settling tables." This has the advantage of settling the largest and "cleanest" starch granules - those that bring the highest price on the market - first. The smaller granules, which may contain clay, settle later and yield a grayish, low-quality flour (Flach 1983). Sunlight is often the sole drying agent for the processed starch.

Water quality is a key factor in the entire procedure: Poor water tends to yield sago starch of lesser quality. The refuse created in the production of sago is only of value if domestic animals are nearby. When this is not the case, the refuse is often simply discarded behind plant buildings, creating a stench that is noticeable at quite some distance (Flach 1983).

Extraction Methods in Different Areas

In New Guinea, good use is made of natural stands of sago palms, as well as planted seedlings and suckers. In the swampy lowlands, the semiwild stands require

only a minimum of pruning. Those who plant and harvest sago palms throughout the year make periodic visits to the various groves to determine the proper time for harvest (Barrau 1960; Ooman 1971; Ohtsuka 1983).

Sago extraction is usually done by extended family groups in New Guinea. The men fell the palm, making the cut approximately 40 to 70 centimeters above the ground. Next, using axes, they remove half of the surface wood (2 to 4 cm thick) in order to expose the pith. While this is going on, the women construct troughs in which the sago starch will be washed out. Once the men have exposed the pith, the women scrape it out of the trunk and pound it into a mass (Barrau 1960; Ohtsuka 1983).

For starch extraction, the Papuans employ an *abol*, a tool made from two hard sticks and a toughened string of cane that is used much like an adze. (In fact, adze-like tools are common throughout New Guinea.) The actual cutting implement is most often made of stone, wood, or sharpened bamboo, although in areas that have contact with Europe, metal piping is frequently employed (Barrau 1960; Ohtsuka 1983).

In New Guinea, as is typical elsewhere, leaves, a trough, and a sieve are used in the kneading and straining process. The starch-bearing liquid is collected in pans made from leaves or leafstalks, then partly dried and wrapped with palm leaves, usually in the shape of a cylinder or a cone. In one study, it was observed that five women, in about 8.5 hours of extracting, sieving, and drying, were able to produce 54.7 kg of sago (Barrau 1960; Ohtsuka 1983).

In Malaysia, the average yield per sago palm has been estimated at between 113 and 295 kg. The fallen palm tree is cut into logs of 120 to 183 cm in length for rasping. The tools used in rasping have evolved from the *palu*, a sharpened bamboo cylinder with a long wooden handle (which caused many leg injuries), to the *garut*, a wooden board with nails, to mechanized scraping machines, introduced in 1931. One such device consists of a spinning metal disc with serrated edges. Kneading is usually done by trampling, and drying takes place in the sun (Knight 1969; Whitmore 1973).

The extraction process that takes place in a factory is quite similar to the more primitive methods already described. In Singapore, for example, an axe is used to remove the bark, and a two-man nail board is employed in the rasping process. Care is taken to process sago trees in the order in which they arrive at the factory, so as to prevent spoilage. The extracted sago is made into blocks, mixed with water, and then blocked again. They dry in the sun, with occasional turning.

Tikopia provides an example of sago extraction in the western Pacific, where the task proceeds during the rainy season because excess water is readily available. Hoops of iron are used to scrape the trunk after the bark is removed; before iron was introduced,

sharp coconut shells were employed. If the work is to be performed in the village instead of in the field, the trunk is cut into sections. Kneading is done in coconut-leaf mesh baskets and the material is then sieved. A trough is filled with the water-and-starch solution and covered with coconut and sago fronds. After the starch has settled, the water is poured off, and the sago is dried and made into flour (Firth 1950).

In South America, where the Warao Indians extract sago from *Manicaria saccifera*, the methods, again, vary only a little from those employed elsewhere. After the tree is felled, the bark is removed, and an adze or hoe (*naburu*) is utilized to rasp the pith. This hoe typically consists of a blade made of *Mauritia* bark, with a handle constructed of rounded wood and a binding consisting of a two-ply cord made from *Mauritia* bast. The trough employed in the process is made from the trunk of the *temiche* palm. After water has been added to the pith and the mixture kneaded through a strainer, a ball of light brown sago is made.

In South America, sago extraction practices may be part of a disappearing tradition, as the starch is slowly giving way to other agricultural staples, even among the tribes who have used it since prehistoric times (Wilbert 1976).

Sago As Food

It is mainly in New Guinea and neighboring islands that *Metroxylon* has been exploited as a food. A typical swamp grove will have approximately 25 palms per acre per year ready for felling. These will yield a total of about 2,837 to 3,972 kg of crude starch, which will provide from 7 to 10 million calories to its consumers. Sago can be used like any other starch, and peoples familiar with it have developed numerous ways of preserving and consuming it (Boulger 1889; Barrau 1960; Flach 1983).

In the swamp areas of New Guinea, where sago is a staple, the average daily ration per person is a little less than a kilogram, with individual consumption ranging from a bit over 0.5 kg to about 1.5 kg per day. Such quantities of sago deliver from 1,700 to about 4,000 daily calories, which the average family in New Guinea devotes 10 days of each month to acquiring (Ooman 1971).

Preservation

Left dry, sago becomes moldy and spoils. But the starch can be stored by simply placing it in a basket, covering it with leaves, and sprinkling water on it from time to time. With moisture, sago ferments and forms lactic acid, which prevents spoiling. If pottery is available, fresh sago is placed in a jar and covered with water (Barrau 1960; Ooman 1971; Flach 1983).

There are, however, methods of storing sago in a dry state. One is to make sago paste into briquettes by dehydrating it rapidly on a surface above a fire. This

method permits the sago to be kept for about one month. Sago can also be dried in the sun, although it is said that this makes it taste "flat" (Barrau 1960; Flach 1983). In general, Papuans tend to think that dried sago loses its flavor.

Supplements

As has been mentioned, nutritional supplements are vital to a diet centering on sago. It must be eaten with some fish or meat (or other whole protein) and with vegetables to provide consumers with a satisfactory intake of the chief nutrients. Thus, in New Guinea, the peoples reliant upon sago, who supplement their diet with fish, hunted game, sago grubs, sago palm heart, leaves, and nuts, probably enjoy a relatively well-balanced diet (Ooman 1971; Dwyer 1985).

Sago Foods

After harvesting, it is common for some of the just-produced sago to be eaten immediately. The women usually prepare it by wrapping a portion in palm leaves or packing it into a section of cane (actually rattan, *Calamus* spp.) and baking it (Ohtsuka 1983). Sometimes, before the sago is baked in a fire, it is mixed with grated coconut or with bean flour (Flach 1983). The remainder of the freshly harvested sago is then wrapped in dry palm fronds to be carried back to the village (Ohtsuka 1983).

The starch is prepared in a number of ways. In areas with pottery, a sago porridge is often served along with condiments, grains, fish, and meat. A biscuit of sago is also made by those who have pottery. In what was Netherlands New Guinea, for example, a sago biscuit was baked in an earthenware mold, which served as the oven.

Areas without pottery will often bake sago paste, rolled in green leaves, in a hot stone oven. This produces a flat cake that often has grated coconut, meat, fish, or greens mixed into it. A cake with grated coconut is called *sago senole*. Sago briquettes, wrapped in sago leaves, are referred to as *sago ega*. *Sago bulu* comes from the cooking of sago paste in green bamboo. A roasted stick of sago paste is called *sago boengkoos*.

In Borneo, sago pellets are used occasionally as a substitute for rice (Barrau 1960). A sago ash may also be produced by burning the wide part of the sago leaf midrib. This can be an important nutritional supplement providing sodium, potassium, calcium, and magnesium.

Pearl sago – another common product from sago starch – is made by pressing wet sago flour through a sieve and then drying it in a pan while stirring continuously. The "pearls" formed are round, and the outer part of the sago pearl gelatinizes to hold them together. Pearl sago is an important ingredient in soups and puddings (Flach 1983). In Sarawak, wet sago flour is mixed with rice polishings and cooked

into pearl form, creating an "artificial rice," certainly a more nutritious food than polished rice. Flach believes that this product has potential as a substitute for rice in Southeast Asia (Flach 1983).

In Tikopia, sago is often made into a flour that is considered a delicacy by those who produce it. On occasion, sago is mixed with other foods to add body, flavor, and softness. Big slabs of sago are also baked for many days in large ovens, and then put aside for times of famine. However, this sago product is considered virtually "unpalatable" by its makers (Firth 1950).

Sago is also employed in foods that are more common in other parts of the world. For example, sago starch can be used in high-fructose syrup as a partial replacement for sucrose (Flach 1983). Sago has also been experimentally added to bread flour. It has been found that adding 10 percent sago to the recipe can improve the quality of the bread produced, although adding more will lower it (Flach 1983).

In addition to the consumption of the palm pith, other parts used as food include the inner shoot of the crown (as fruit or snack), sap from the male inflorescence (boiled into palm sugar, fermented as vinegar or distilled spirit), and the inner kernel (cooked in syrup as a dessert) (Lie 1980). Overall, the uses of sago are as varied as those of other starches.

H. Micheal Tarver
Allan W. Austin

Bibliography

- Avé, J. B. 1977. Sago in insular Southeast Asia: Historical aspects and contemporary uses. In *Sago-76: Papers of the first international sago symposium*, ed. Koolin Tan, 21–30. Kuala Lumpur.
- Barrau, Jacques. 1960. The sago palms and other food plants of marsh dwellers in the South Pacific islands. *Economic Botany* 13: 151–62.
- Boulger, G. S. 1889. *The uses of plants: A manual of economic botany*. London.
- Dwyer, Peter D. 1985. Choice and constraint in a Papuan New Guinea food quest. *Human Ecology* 13: 49–70.
- Firth, Raymond. 1950. Economics and ritual in sago extraction in Tikopia. *Mankind* 4: 131–43.
- Flach, M. 1983. *The sago palm*. Rome.
- Isaac, Erich. 1970. *Geography of domestication*. Englewood Cliffs, N.J.
- Knight, James Wilfred. 1969. *The starch industry*. Oxford.
- Lie, Goan-Hong. 1980. The comparative nutritional roles of sago and cassava in Indonesia. In *Sago: The equatorial swamp as a natural resource*, ed. W. R. Stanton, and M. Flach, 43–55. The Hague.
- McCurrach, James C. 1960. *Palms of the world*. New York.
- Moore, Harold E., Jr. 1973. Palms in the tropical forest ecosystems of Africa and South America. In *Tropical forest ecosystems in Africa and South America: A comparative review*, ed. Betty J. Meggers, Edward S.

- Ayensu, and W. Donald Duckworth, 63–88. Washington, D.C.
- Morris, H. S. 1974. In the wake of mechanization: Sago and society in Sarawak. In *Social organization and the applications of anthropology*, ed. Robert J. Smith, 271–301. Ithaca, N.Y.
- Murai, Mary, Florence Pen, and Carey D. Miller. 1958. *Some tropical South Pacific island foods: Description, history, use, composition, and nutritive value*. Honolulu.
- Ohtsuka, Ryutaro. 1983. *Oriomo Papuans: Ecology of sago-eaters in lowland Papua*. Tokyo.
- Ooman, H. A. P. C. 1971. Ecology of human nutrition in New Guinea: Evaluation of subsistence patterns. *Ecology of Food and Nutrition* 1: 1–16.
- Peters, F. E. 1957. *Chemical composition of South Pacific foods*. Noumea, New Caledonia.
- Platt, B. S. 1977. *Table of representative values of foods commonly used in tropical countries*. London.
- Rhoads, James W. 1982. Sago palm management in Melanesia: An alternative perspective. *Archaeology in Oceania* 17: 20–4ff.
- Ruddle, Kenneth. 1977. Sago in the new world. In *Sago-76: Papers of the first international sago symposium*, ed. Koonlin Tan, 53–64. Kuala Lumpur.
- Ruddle, Kenneth, Dennis Johnson, Patricia K. Townsend, and John D. Rees. 1978. *Palm sago: A tropical starch from marginal lands*. Honolulu.
- Stanton, W. R., and M. Flach. 1980. *Sago: The equatorial swamp as a natural resource*. The Hague.
- Tan, Koolin. 1977. *Sago-76: Papers on the first international sago symposium*. Kuala Lumpur.
1980. Logging the swamp for food. In *Sago: The equatorial swamp as a national resource*, ed. W. R. Stanton and M. Flach, 13–34. The Hague.
- Whitmore, Timothy. C. 1973. *Palms of Malaya*. Kuala Lumpur.
- Wilbert, Johannes. 1976. *Manicaria saccifera* and its cultural significance among the Warao Indians of Venezuela. *Botanical Museum Leaflets* 24: 275–335.

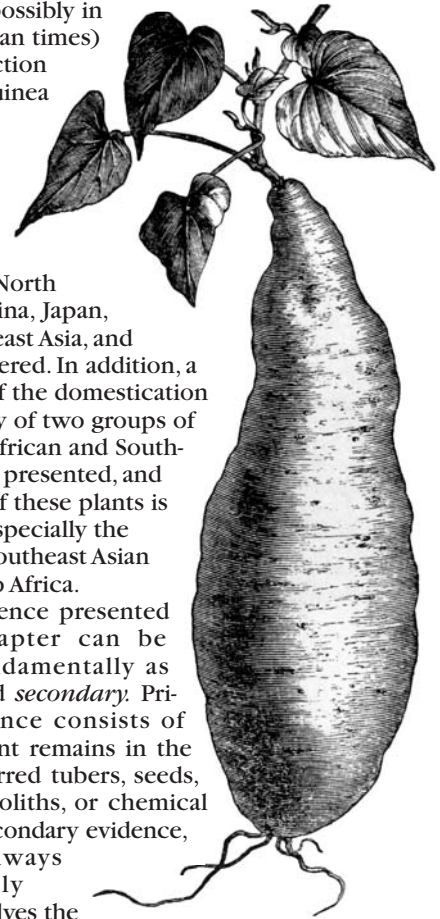
II.B.5 🍠 Sweet Potatoes and Yams

The sweet potato (*Ipomoea batatas*, Lam.) and the yams (genus *Dioscorea*) are root crops that today nurture millions of people within the world's tropics. Moreover, they are plants whose origin and dispersals may help in an understanding of how humans manipulated and changed specific types of plants to bring them under cultivation. Finally, these cultivars are important as case studies in the diffusion of plant species as they moved around the world through contacts between different human populations.

This chapter reviews the questions surrounding the early dispersals of these plants, in the case of the sweet potato from the New World to the Old, and in the case of yams their transfers within the Old World. In so doing, the sweet potato's spread into Polynesia before European contact is documented, and the issue of its penetration into

Melanesia (possibly in pre-Columbian times) and introduction into New Guinea is explored. Finally, the post-Columbian spread of the sweet potato into North America, China, Japan, India, Southeast Asia, and Africa is covered. In addition, a discussion of the domestication and antiquity of two groups of yams, West African and Southeast Asian, is presented, and the spread of these plants is examined, especially the transfer of Southeast Asian varieties into Africa.

The evidence presented in this chapter can be viewed fundamentally as *primary* and *secondary*. Primary evidence consists of physical plant remains in the form of charred tubers, seeds, pollen, phytoliths, or chemical residuals. Secondary evidence, which is always significantly weaker, involves the use of historical documents (dependent on the reliability of the observer), historical linguistics (often impossible to date), stylistically dated pictorial representations (subject to ambiguities of abstract representation), remnant terracing, ditches or irrigation systems (we cannot know which plants were grown), tools (not plant specific), and the modern distribution of these plants and their wild relatives (whose antiquity is unknown).



Sweet potato

The Sweet Potato

In general, studies of the origin of domesticated plants have first attempted to establish genetic relationships between these plants and some wild ancestor. In the case of the sweet potato, all evidence employed by previous archaeological, linguistic, and historical investigators establishes its origins in the New World.

The remains of tubers excavated from a number of archaeological sites in Peru provide the most persuasive evidence for this conclusion. The oldest evidence discovered to date is from the central coast region at Chilca Canyon where excavated caves, called Tres

Ventanas, yielded remains of potato (*Solanum* sp.), of jicama (*Achirbizus tuberosus*), and of sweet potato (*Ipomoea batatas*) (Engel 1970: 56). The tubers were recovered from all levels, including a level in Cave 1 dated to around 8080 B.C.

These plants were identified by Douglas Yen who could not determine if they were wild or cultivated species, although he observed that wild tuber-bearing sweet potatoes today are unknown (Yen 1976: 43). Whether they ever existed is another matter, but at least the findings in these cases suggest the consumption of "wild" sweet potato at 8000 B.C. in Peru, or more radically, a domesticated variety (Hawkes 1989: 488). If the latter is the case, sweet potatoes would be very ancient in the New World, raising the possibility that perhaps sweet potatoes were the earliest major crop plant anywhere in the world.

Sweet potato remains were also present at a Pre-ceramic site, Huaynuma, dating around 2000 B.C., and at an Initial Period site, Pampa de las Llamas-Moxeke in the Casma Valley, dating from around 1800 to 1500 B.C. (Ugent, Pozorski, and Pozorski 1981: 401-15). In addition, remains have been found at the Early Ceramic Tortugas site in the Casma Valley (Ugent and Peterson 1988: 5).

Still other sweet potato remains in Peru were discovered at Ventanilla, dating from around 2000 to 1200 B.C., the Chillon Valley (Patterson and Lanning 1964: 114), the central coast in the third phase of the Ancon sequence dating 1400 to 1300 B.C. (Patterson and Moseley 1968: 120), and the third Colinas phase dating 1300 to 1175 B.C. (Patterson and Moseley 1968: 121).

Thus, archaeological evidence gives a date of *at least* 2000 B.C. for the presence of the domesticated sweet potato in the New World, while suggesting a possible domesticated form as early as 8000 B.C.

The Botanical Data

In the past, a precise identification of the ancestor of the sweet potato was hampered by a lack of taxonomic concordance. However, Daniel Austin's proposal (1978: 114-29) of a Batatas complex with eleven closely related species represents a significant revision. With other species it includes *I. trifida*, which is often identified as the key species for the origin of *Ipomoea batatas* (Orjeda, Freyre, and Iwanaga 1990: 462-7). In fact, an *I. trifida* complex, to include all plants of the section Batatas that could cross with *I. batatas*, has been proposed (Kobayashi 1984: 561-9).

In the early 1960s, a wild Mexican species of *Ipomoea* (No. K-123) that was cross-compatible and easily hybridized in reciprocal crosses was identified (Nishiyama 1961: 138, 1963: 119-28). Though it resembled the sweet potato, it lacked domesticated traits (Nishiyama 1961: 138, 1963: 119-28). Cytological studies showed K-123 had a chromosome number

($n = 45$) similar to the sweet potato, and it was identified as *I. trifida* (Nishiyama, Miyazakim, and Sakamoto 1975: 197-208). One concern with K-123 was that it might be a feral sweet potato (Austin 1983: 15-25). But other research proposed that *I. leucantha* was crossed with a tetraploid *I. littoralis* to produce the hexaploid *I. trifida*, with the sweet potato selected from this hexaploid (Martin and Jones 1986: 320).

Critical to the debate is the discovery of the natural production of $2n$ pollen in 1 percent of diploid *I. trifida* (Orjeda, Freyre, and Iwanaga 1990: 462-7). The $2n$ pollen is larger than the n pollen, and the diploid populations of *I. trifida* exhibit gene flow between diploids and tetraploids (Orjeda, Freyre, and Iwanaga 1990: 462). Fundamentally, crosses between n and $2n$ pollens make various $2\times$, $3\times$, $4\times$, $5\times$, and $6\times$ combinations of the *I. trifida* complex possible and could result in $6\times$ combinations leading to the sweet potato (Orjeda, Freyre, and Iwanaga 1990: 466). The plants exhibiting this feature come predominantly from Colombia in northwest South America (Orjeda, Freyre, and Iwanaga 1990: 463).

While this new evidence shows how the sweet potato could have arisen from *I. trifida*, the present evidence fails to account for the enlarged storage tuber, nonclimbing vines, red periderm color, and orange roots of the sweet potato (Martin and Jones 1986: 322). It is interesting to note the report of Masashi Kobayashi (1984: 565) that some Colombian varieties of *I. trifida* produce tuberous roots at high elevations. Typically, these plants are found at much lower elevations of between 5 to 20 meters (m) above sea level in Colombia, although some occur at about 1000 m.

Given these observations, it should be mentioned that it is often assumed, after a new species arose through sexual reproduction, that it reached a static stage and that vegetative multiplication has nothing to do with the origin of new forms (Sharma and Sharma 1957: 629). In the various researches into the origin of the sweet potato, investigators have assumed the species arose through sexual reproduction, but karyotypic alterations in somatic cells are common in vegetative reproducing plants, and speciation can occur at those points (Sharma and Sharma 1957: 629). The sweet potato is known to send out roots from the nodes, which will bear small potatoes (Price 1896: 1); therefore, it is possible that karyotypic alterations occurred in the daughter forms, giving rise to new forms, as in some species of *Dioscorea*. This is important because spontaneous mutations occur quite often, and the sweet potato mutates easily using gamma and X rays (Broertjes and van Harten 1978: 70).

In summary, 20 years ago, the singling out of any one species of *Ipomoea* as the ancestral form of the sweet potato represented no more than an educated guess (O'Brien 1972: 343; Yen 1974: 161-70). But today the evidence points strongly to the *I. trifida* complex of plants found in northwestern South America.

Dispersal

Present evidence shows that the sweet potato was introduced into Europe, Asia, Africa, and Australia after Christopher Columbus reached the New World. There is no data to indicate that the plant was known to the ancient civilizations of China, Egypt, Babylon, Persia, Rome, Greece, or India (Cooley 1951: 378), but there is evidence of a pre-Columbian introduction into Oceania. Therefore, in this section, a pre-Columbian and a post-Columbian spread of the sweet potato is outlined, starting with the post-Columbian transfer.

The Post-Columbian Spread

Europe. The sweet potato was introduced into Europe via Spain at the end of the fifteenth century by Christopher Columbus and Gonzolo Fernandez de Oviedo (de Candolle 1959: 55). From this beginning, it spread to the rest of Europe and was called *batata* and *padada* (Cooley 1951: 379).

United States. The sweet potato was brought to Londonderry in New Hampshire by the Scotch-Irish in 1719 (Safford 1925: 223). Yet sweet potatoes are mentioned as growing in Virginia in 1648, and perhaps as early as 1610. Further, they are mentioned in 1781 by Thomas Jefferson (Hedrick 1919: 315). They were also reportedly introduced into New England in 1764, and in 1773 the Indians in the South were reported to be growing them (Hedrick 1919: 315-16).

China. Ping-Ti Ho writes that two theories exist for the introduction of sweet potatoes into China. The first involves an overseas merchant who brought the plants from Luzon, which were given to the governor of Fukien in 1594 to alleviate a famine (Ho 1955: 193). The second claim suggests that sweet potatoes arrived via the southern port of Chang-chou, but no specific date of this alleged introduction is known (Ho 1955: 193). Ping-Ti Ho indicates that whereas the former story may be true, the sweet potato was already in China by 1594, having been observed by 1563 in the western prefecture Ta-li, near Burma (Ho 1955: 193-4). Ho concludes that the introduction could have been either overland or by sea via India and Burma, well before the generally accepted date of 1594.

Japan. An English factory at Hirado was allegedly responsible for first introducing the sweet potato to Japan about 1615. It did not, however, "catch on," and the plant was reintroduced from China in about 1674 to stave off a famine (Simon 1914: 716, 723-4).

India and Southeast Asia. The sweet potato was introduced into India by the Portuguese who brought it to Macão via Brazil (Zavala 1964: 217). Moreover, a Portuguese influence in the spread of the plant to Ambon, Timor, and parts of the northern Moluccas is indicated linguistically, since names for

the plant are variations of the word *batata* (Conklin 1963: 132).

In Malaysia the sweet potato is called Spanish tuber (Conklin 1963: 132). In the Philippines it is called *camote* (Merrill 1954: 161-384), whereas in Guam it is called both *camote* and *batat* (Hornell 1946: 41-62; Conklin 1963: 129-36). In the Moluccas and on Cebu it is called *batat* (Merrill 1954: 161-384). The names themselves indicate the source of the plant, since the Portuguese used the Arawak term (*batata*), whereas the Spanish employed the Nahuatl one (*camote*).

Africa. Harold Conklin (1963: 129-36) reports that the terms for the sweet potato take the form "*batata*, *tata*, *mbatata*, and the like" from much of the African continent south of the Sahara. The other most widespread term is *bombe*, *bambai*, *bambaira*, or *bangbe*, with the Indian trade center of Bombay the etymological source of the word and its pronunciation. Names for sweet potato not falling into these categories have a small geographic distribution or are found only in closely related languages, but as lexical entries, they refer to indigenous yams or other root crops. From these data Conklin concludes that the sweet potato was introduced into Africa by the Portuguese from Brazil, probably early in the slave-trade period of the sixteenth century (Conklin 1963: 129-36). In addition, its introduction into West Africa, specifically Angola, probably coincided with Paulo Dias de Novais's charter of colonization in 1571, which included provisions for peasant families from Portugal with "all the seeds and plants which they can take from this kingdom and from the island of São Tomé" (Boxer 1969: 30).

The Portuguese ports of Mozambique probably saw the introduction of the *batata* from India as well, although the presence of the word *batata* cannot always be taken as indicative of a Portuguese association. Their language became the *lingua franca* along the coasts of Africa and Asia (Boxer 1969: 55).

The word *bambai*, by contrast, is obviously linked to the city of Bombay, but that port was not significant in the India trade network until the British acquired it in 1662. Consequently, for the word Bombay to be associated with the sweet potato in Africa suggests either a British connection or possibly Indian clerks and merchants involved with British colonization. Consequently, Bombay's involvement in introducing the sweet potato to Africa could have been as early as the last quarter of the seventeenth century or as late as the nineteenth (O'Brien 1972: 347).

Thus, the evidence seems to suggest that the sweet potato was introduced into Africa by the Portuguese from Brazil and Lisbon in the sixteenth century. A later spread of the plant possibly occurred via British influence in the late seventeenth through nineteenth centuries.

The Pre-Columbian Spread

Polynesia. The traditional view is that the sweet potato was introduced into Polynesia by the Spanish, who brought it to the Philippines in 1521 (Dixon 1932: 41). Moreover, its name in the Philippines, *camote*, is generically related to the Nahuatl *camotl* and *camotilli* (Merrill 1954: 317-18). Another point linked to this theory is that the earliest European Pacific explorers, Alvaro de Mendaña and Pedro Fernández de Quiros, did not mention the plant by 1606, although they had visited the Marquesas, Santa Cruz, the Solomons, and New Hebrides (Yen 1973: 32-43).

Yet scholars have also argued that the sweet potato was introduced into Polynesia long before Ferdinand Magellan's 1521 voyage, based on the fact that sweet potatoes were found to be a major part of the economies of the islands located at the points defining the triangle of Polynesia, at the time of their discovery - these being New Zealand in 1769, Hawaii in 1778, and Easter Island in 1722 (Dixon 1932: 45).

Further support for the antiquity of the sweet potato in Polynesia has to do with the very large numbers of varieties found in the South Seas: 48 in New Zealand (Colenso 1880: 31-5), 24 in Hawaii (Handy 1940: 133-5), 16 in the Cook Islands, and 22 in New Guinea (Yen 1961: 368, 371).

Twenty years ago the best evidence to document an early introduction of the sweet potato to Polynesia was historical linguistics that had reconstructed the word for sweet potato in Proto-Polynesian to *kumala* (O'Brien 1972: 349-50). Over the years, other scholars have scrutinized the proposed antiquity of the word for sweet potato, and believe that a Proto-Polynesian origin of the word is plausible (Pawley and Green 1971: 1-35; Biggs 1972: 143-52; Clark 1979: 267-8).

Such linguistic evidence establishes a base line for the antiquity of the sweet potato in Polynesia, and when combined with archaeological information about the peopling of the Pacific, it is possible to hypothesize the approximate time of entry of the plant to the region. Jesse Jennings (1979: 3) suggests a Polynesian presence on Tonga and Samoa around 1100 and 1000 B.C., respectively, with an initial thrust east into the Marquesas by A.D. 300. This early appearance was probably associated with the Lapita penetration of western Polynesia at around 1500 B.C. from Melanesia (Bellwood 1978: 53).

And finally, in the past ten years, another line of secondary evidence has been investigated in New Zealand, where prehistoric storage facilities and man-made soils had been discovered (Leach 1979: 241-8, 1984, 1987: 85-94).

However, much *primary evidence* also exists to indicate a pre-Columbian introduction of the sweet potato into Polynesia.

Hawaiian Islands. Archaeological evidence for antiquity of the sweet potato in Hawaii has been found in the form of a carbonized tuber from a fireplace

within a "middle" phase domestic structure at Lapakahi. The fireplace is dated A.D. 1655 $\pm$ 90 with a corrected date of A.D. 1635 $\pm$ 90 or A.D. 1515 $\pm$ 90, giving a range of A.D. 1425 to 1765 (Rosendahl and Yen 1971: 381-3). James Cook visited Hawaii in 1778, and so it would seem that this tuber was incinerated at least 13 years prior to his arrival and potentially as many as 263 years before.

Easter Island. The sweet potato was the major crop plant on Easter Island when it was discovered by Jacob Roggeveen in 1722. Charred remains of the plant were recovered there from a fireplace dating A.D. 1526 $\pm$ 100 (Skjolsvold 1961: 297, 303). This gives a range between A.D. 1426 to 1626, making the plant remains pre-European by at least 96 years.

New Zealand. In New Zealand, Maori traditions of reconstructing lineage genealogies back to A.D. 1350 recount the arrival of some mysterious "fleet" with the sweet potato and other domesticated plants and animals aboard (Golson 1959: 29-74). Archaeological evidence for the early presence of the sweet potato may exist in the form of ancient storage pits. Jack Golson, for example, (1959: 45) has argued that pits excavated at the fourteenth-century Sarahs Gully site, may have been storage pits for sweet potatoes (*kumara*). To R. Garry Law (1969: 245) as well, sites like Kaupokonui, Moturua Island, Skippers Ridge, and Sarahs Gully give evidence of widespread *kumara* agriculture by A.D. 1300.

Primary archaeological evidence was furnished by Helen Leach (1987: 85) with her discovery of charred sweet potato tubers in a burned pit at a *pa* site from (N15/44) the Bay of Islands. Locally called "Haratua's pa," the site is prehistoric, as are the charred sweet potatoes, a point that seems to confirm that these pits were used to store them (Sutton 1984: 33-5).

In addition to the charred tubers at the Bay of Islands, a single charred tuber was discovered at Waioneke, South Kaipara (Leach 1987: 85), a "classic" Maori site 100 to 300 years old (Rosendahl and Yen 1971: 380). Helen Leach (personal communication, letter dated 13 Feb. 1992) notes that no European artifacts were present, and therefore she considers "these *kumara* pre-European in origin."

Central Polynesia. The most exciting recent evidence dealing with the antiquity of the sweet potato in Polynesia is the discovery of charred *kumara* tubers at site MAN-44 on Margaia Island in the Cook Island group dated at A.D. 1000 (Hather and Kirch 1991: 887-93). The presence of charred remains this early seems to establish beyond doubt a pre-Columbian introduction into Polynesia.

Micronesia. In the Carolinas, infrared spectroscopy analyses of organic residues found on pottery has documented the presence of the sweet potato (and

taro) at the Rungruw site in the southern part of Yap at about A.D. 50 (Hill and Evans 1989: 419–25). The presence of rice and banana at about 200 B.C. at the same site was also established (Hill and Evans 1989: 419–25). Yap and the Carolinas are near the northern fringe of Melanesia.

Melanesia. The spread of the sweet potato into Melanesia appears to be the result of Polynesian and European introduction, with the former probably ancient. When the Solomons were discovered, there was no mention of the plant, although taro and yams were reported (Mendana 1901: 212). Because Polynesians were present in the Solomons, it is possible that they brought the plant, since the word *kumala* is found in Melanesian pidgin on the islands (O'Brien 1972: 356).

The term *kumala* is used in New Caledonia and may be pre-European (Hollyman 1959: 368). The sweet potato was in this area in 1793 (Hollyman 1959: 368), and was grown, in lesser quantities than present, in precontact Fiji (Frazer 1964: 148). Finally, there is evidence of the plant's presence in the New Hebrides at the time of discovery (Dixon 1932: 42–3).

New Guinea. New Guinea is the one region of Oceania where the sweet potato is of profound economic importance today. It is more widely grown in the western part of the island than in the east (Damm 1961: 209) and is of great importance in the highlands of Irian New Guinea (Damm 1961: 209). Among the inhabitants of Wantoat Valley in highland Papua, the sweet potato is the only important cultivated food (Damm 1961: 212–3).

Dating the introduction of the sweet potato into New Guinea, however, is a problem. Some historical data point to a late entry. For example, it was introduced into the Morehead district by missionaries (Damm 1961: 210) and was even more recently introduced to the Frederick-Hendrick Island region (Serpenti 1965: 38). Moreover, a survey of plants and animals in 1825–6 revealed no sweet potatoes on the islands west of New Guinea, on the island of Lakor, on the Arru and Tenimber islands, and the southwest coast of Dutch New Guinea (Kolff 1840). The plant ecologist L. J. Brass has suggested that the sweet potato came from the west some 300 years ago, carried by birds of paradise and by hunters and traders in the Solomons region (Watson 1965: 439), which may point to a European introduction.

Introduction to the South Pacific

The primary evidence available today suggests that the sweet potato had a prehistoric introduction into Polynesia and Micronesia at around the time of Christ, while the linguistic evidence points to its presence during Proto-Polynesian times. If Proto-Polynesian was the language of the Lapita culture populations, then the sweet potato was present in Oceania possibly as

early as 1500 B.C. Given these new data, the next question must be about the mechanism that facilitated its transfer into Oceania at this early date, since the plant is definitely a New World species.

In attempting to answer that question, a number of researchers over the years have been struck by the similarity between the Polynesian words for sweet potato (*kumala*, *kumara*) and the word *cumara* found in some Quechua language dictionaries (Brand 1971: 343–65). This, in turn, has led to the suggestion that the sweet potato came to Polynesia from Peru, with Quechua speakers playing some role.

Alternately, Donald Brand (1971: 343–65) argues that the word was Polynesian and introduced into the Andes by the Spanish. He notes that archaeologists, historians, and philologists consider coastal Ecuador, Peru, and Chile to have been occupied by non-Quechuan and non-Aymaran people until shortly before the arrival of the Spanish. The languages spoken would have been Sek, Yungan, and Chibchan, and their terms for sweet potato were *chapru*, *open*, and *unt*. The Quechua word is *apichu* and is reported along with the words *batatas*, *ajes*, and *camotes* in the literature and dictionaries of Peru, whereas the word *cumara* is found only in dictionaries, and *cumar* proper occurs only in the Chichasuyo dialect of Quechua (Brand 1971: 361–2).

If it is true that the Spanish introduced the word, then one need not explain its presence in Polynesia as the result of people going or coming from the New World. And if the word *kumala* is Proto-Polynesia, then the term was created by the Polynesians for a plant in their cosmos.

But this still leaves the question of how it might have entered that cosmos. Since the tuber cannot float at all, let alone the thousands of miles separating Oceania and northwest South America, only two explanations appear possible: Transference was accomplished by either a human or a nonhuman agent.

A human agency might have involved a vessel with sweet potatoes aboard drifting from the New World and being cast upon one of the islands of western Polynesia, like Samoa. If any members of the crew survived, they might well have passed along the South American name for the plant. On the other hand, if an empty vessel reached an inhabited island, it would have been examined along with its cargo, and the sweet potato, looking a great deal like a yam, might have been treated like one until its particular features were known. Finally, during a vessel's long drift, rain water might have accumulated within it, in which case the tubers would have begun to grow, taking hold first in the vessel and then in the soil of some uninhabited island, ultimately becoming feral. Later people finding both the island and the plants would have redomesticated and named them.

An alternative possibility would be transfer by a natural agent. Sweet potato tubers cannot float, but its

seeds are more mobile, making birds a likely vehicle. Indeed, Douglas Yen (1960: 373) has suggested the possibility of birds as an agent, and Ralph Bulmer (1966: 178–80) has examined the role of birds in introducing new varieties of sweet potato into gardens in New Guinea by dropping seeds. Bulmer observed that the golden plover, a bird that ranges over Polynesia, is a casual visitor to the west coast of the Americas as far south as Chile. These birds are strong fliers and could have carried the small, hard sweet potato seeds either in their digestive tracts or adhering to mud on their feet.

Another potential nonhuman agent was proposed by J. W. Purseglove (1965: 382–3), who noted that some species of *Ipomoea* are strand plants and are distributed by sea. He points out that dried sweet potato capsules with several seeds can float. Because the Polynesian sweet potatoes are very distinctive, he suggests that this distinctiveness is the predictable result of an introduction by a few seeds. Purseglove also observes that introduced crop plants have a considerable advantage if major pests and diseases have not been transferred.

At present, all of these scenarios are only speculative, but an accidental introduction would explain how the plant reached the area early, and yet account for the absence of other useful New World products (manioc, maize, and so forth), which might have been transferred if any sustained exchange between living people had been involved.

The Yams

Although a number of wild members of *Dioscorea* are edible, there are four domesticated yams that are important to agricultural development: *D. alata* and *D. esculenta* from Southeast Asia, and *D. rotundata* and *D. cayenensis* from West Africa. A fifth domesticated yam, *D. trifida*, is found in the New World (Hawkes 1989: 489), but was not especially significant because of the presence of the sweet potato, manioc (cassava), and potato (Coursey 1975: 194) and, thus, has not been a specific focus of research.

The Southeast Asian varieties are interesting because aspects of their spread into Polynesia can be linked to the spread of the sweet potato, whereas African varieties are significant for the role they played in the development of the kingdoms of West Africa. Like the sweet potato, there is no evidence of the use of yams in classical antiquity, but historical data point to their presence in China in the third century A.D. and in India by A.D. 600 (Coursey 1967: 13–14).

The Botanical Data

The family *Dioscoreaceae* has hundreds of species, and the *Dioscorea* is its largest genus. In general, the New World members have chromosome numbers that are multiples of nine and the Old World species

multiples of ten (Ayensu and Coursey 1972: 304). The section including the food yams typically has $2n = 40$, but higher degrees of polyploidy do occur (Coursey 1967: 43–4). For example, *D. alata* has $2n = 30$ to 80; *D. esculenta* has $2n = 40, 60, 90, 100$; *D. rotundata* has $2n = 40$; and *D. cayenensis* has $2n = 40, 60$, and 140. *D. trifida*, the New World domesticated yam, has $2n = 54, 72$, and 81 (Coursey 1976a: 71).

The two yams domesticated in Southeast Asia have been major constituents (along with taro, plantains, bananas and breadfruit) of root crop agriculture in the region, and throughout Oceania before European contact. According to D. G. Coursey (1967: 45), *D. alata*, or the “greater yam,” is native to Southeast Asia and developed from either *D. hamiltonii* Hook. or *D. persimilis* Prain et Burk. It is unknown in the wild state and today is the major yam grown throughout the world (Coursey 1967: 45–6). *D. esculenta*, or the “lesser yam,” is a native of Indochina, has smaller tubers than the “greater yam” (Coursey 1967: 51–2), and occurs in both wild and domesticated forms (Coursey 1976a: 71). The two native African yams, probably ennobled in the yam belt of West Africa, are *D. rotundata* Poir., the white Guinea yam, and *D. cayenensis* Lam., the yellow Guinea yam (Coursey 1967: 11, 48, 58). The most prominent English-speaking scholars to work on the genus *Dioscorea* have been I. H. Burkill and D. G. Coursey (1976b, 1980; see also Coursey 1967: 28 for further discussion). Indeed, Coursey, the preeminent yam ethnobotanist, has developed a detailed theory of their domestication and antiquity in Africa.

Nonetheless, African yams have not received the attention of plant scientists that they need and deserve, especially in terms of cytological research and breeding programs (Broertjes and van Harten 1978: 73–4). This omission is particularly regrettable in light of their ancient importance, but is doubtless the result of yams being displaced by New World cultivars like maize, sweet potato, and manioc in many parts of Africa.

The lack of botanical research, however, allows plenty of room for controversy. For example, some botanists separate West African yams into two species, whereas others argue that there are insufficient criteria (basically, tuber flesh color) to separate them. They suggest the existence of a *D. cayenensis-rotundata* complex under the rubric of one species, *D. cayenensis* (Miege 1982: 377–83).

D. G. Coursey, as mentioned, identifies the two yams as *D. rotundata* Poir., and *D. cayenensis* Lam. (Coursey 1967: 11, 48, 58). He suggests that the former is unknown in the wild, being a true cultigen, and it may have developed from *D. praebensilis* Benth. (Coursey 1967: 59). The latter, however, is found in both a wild and domesticated condition (Coursey 1967: 48), which may indicate that the wild *D. cayenensis* is the ancestor of the domesticated *D. cayenensis*.

J. Miege (1982: 380–1) states that *D. cayenensis* is a complex cultigen most probably made up of several wild species: *D. praehensilis* Benth. for the forest varieties; and *D. sagittifolia* Pax., *D. lecardii* De Wild., *D. liebrechtsiana* De Wild., and *D. abyssinica* Hochst. ex. Kunth for the savanna and preforest types. An implication of the argument that these two domesticated species are but subspecies of *D. cayenensis* is that both the white and yellow Guinea yams could have risen from wild forms of *D. cayenensis*.

Clearly, such uncertainties will only be resolved by concerted research focused not only upon taxonomic issues but especially on cytological ones. A whole series of breeding and cross-breeding studies are essential, and it would be particularly useful to determine whether *Dioscorea* polyploidy is related to $2n$ pollen as it is in *Ipomoea*.

Transformation and Dispersal

As we noted, the four major domesticated yams come from Southeast Asia and West Africa, respectively. This section examines data, primary and secondary, for their antiquity and their movement throughout the world.

Southeast Asia. In the late 1960s, charred and uncharred botanical material was recovered from excavations at Spirit Cave in Thailand. It was associated with the Hoabinhian complex, dated to around 10,000 to 7000 B.C., and was used to argue for the early development of agriculture in Southeast Asia (Gorman 1969, 1970). Later, however, these materials were critically reexamined by Yen (1977: 567–99), who concluded that most of the remains were not domesticates. Yen thought that early yam domestication could not be inferred from these remains, but that it was probably reasonable to suspect that wild yam was being eaten at that time (Yen 1977: 595).

The fundamental evidence for the antiquity of domesticated Southeast Asian yams and other cultivars is linguistic and lies within words for the whole assemblage of plants and animals making up Southeast Asian root crop agriculture. Robert Blust, a linguist, notes (1976: 36) that Proto-Austronesian speakers had pigs, fowl, and dogs and cultivated a variety of root and tree crops including taro, yams, sago, breadfruit, sugarcane, and banana (Blust 1976: Table II.B.6.1). The linguist Ross Clark reports that words for all the crop plants important in Polynesia horticulture – yam, taro, bananas, sugarcane, and sweet potato – reconstruct to Proto-Polynesian (Clark 1979: 267–8). In relation to this, it should be mentioned that a Lapita site on Fiji, dating between 1150 to 600 B.C., has primary evidence for aspects of this economy in the form of bones of dogs, chickens, and pigs (Hunt 1981: 260).

Helen Leach (1984: 20–1) believes that a series of 21 roundish holes about 35 centimeters (cm) in diam-

eter and some 60 cm deep excavated within a 33 square meter area at Mt. Olo in Samoa implies yam cultivation, for she reports that large yams on Fiji and in other parts of Melanesia are planted in individual holes 60 cm in diameter and 45 cm deep. She also argues for the antiquity of root crop agriculture at Pailiser Bay in New Zealand through indirect evidence such as storage pits, garden boundaries, old ditches, and “made-soils” (Leach 1984: 35–41). Susan Bulmer (1989: 688–705) makes these same points, but emphasizes the importance of forest clearance, which in New Zealand appears as early as A.D. 400. Indeed, the antiquity of root crop agriculture in New Guinea is documented by this same type of indirect evidence, and Jack Golson outlines a five-phase model of agricultural development and intensification based upon a whole series of field drainage systems that can be dated as early as 7000 B.C. (Golson 1977: 601–38).

In sum, the evidence, though more indirect than direct, supports the notion that the domestication of the Southeast Asian yams, *D. alata* and *D. esculenta*, is very ancient, maybe as early as 4500 B.C. This being the case, what of their dispersal?

The first dispersal is clearly associated with its transfer by Proto-Austronesian-speaking peoples throughout the Southeast Asian tropical world. However, the diffusion of these people is in some dispute. For example, Peter Bellwood (1985: 109) argues that the original Pre-Austronesians were located in Taiwan, whence they moved to the Philippines and from there to parts of Indonesia like Borneo, Java, Sumatra, and Malaya, then into the Moluccas, and finally into New Guinea and Oceania (Melanesia, Micronesia, and Polynesia). But Wilhelm Solheim (1988: 80–2) suggests that Pre-Austronesians developed around 5000 B.C. in Mindanao and the northeast Indonesia regions. He argues against Taiwan as a homeland because of the difficulties posed by winds and currents for sailing south to the Philippines. William Meacham (1988: 94–5), however, considers the languages of south China to have been Mon-Khmer, not Austronesian, and argues that these people could not have migrated to Taiwan and from there south into the Philippines. Rather, Meacham suggests, the homeland of the Proto-Austronesians is somewhere in the triangle defined by Taiwan, Sumatra, and Timor, basically encompassing modern Indonesia.

Regardless of which theory of a Proto-Austronesian homeland is correct, once the Proto-Oceanic languages of that family began to differentiate, they also began to provide linguistic evidence for yams and other cultivars. Thus, yams were in Melanesia by 2000 B.C., in Micronesia by 2000 to 1000 B.C., and in eastern Polynesia by 1500 B.C. The bulk of western Polynesia received yam horticulture (depending on when a specific island group was occupied) sometime between A.D. 1 and 1000 (Bellwood 1985: 121).

In addition, the transfer of Southeast Asian yams with Austronesian speakers to regions outside this

early core area is documented. They were present in China in the third century A.D. and in India by A.D. 600 (Coursey 1967: 13-14). To the west, yams were introduced (principally *D. alata*) into Madagascar, probably between the eleventh and fifteenth centuries A.D. By the end of the sixteenth century, *D. alata* was grown in West Africa, from whence it was transferred to the New World by a Dutch slaver in 1591 (Coursey 1967: 15-6).

Africa. The student of African yams, D. G. Coursey, argues (1967: 13; 1975: 203; 1976a: 72) that the use of *D. cayenensis* and *D. rotundata* is ancient, and he proposes the following scenario for the process of yam domestication in West Africa (1980: 82-5). He suggests:

1. that hunter-gatherers, before 60,000 B.P. (before the present), utilized many species of wild yam;
2. that the Sangoan and Lupemban Paleolithic stone industries, c. 45,000 to 15,000 B.P., developed hoes or picks to excavate hypogeous plants, including the yams, and at this time started to develop ritual concepts and sanctions to protect these and other plants;
3. that sometime around 11,000 B.P., with the contraction of West Africa forest and savanna environments and appearance of proto-Negro people, microlithic industries developed which point to new human/environment interactions; these interactions involved selection and protection of favored species, particularly nontoxic yams; and this greater control led to population increases, movement into forest environments and a planting of wild plants – a “protoculture” with a final result being the understanding that one could replant stored tubers;
4. that by 5,000 to 4,000 B.P. Neolithic grain-crop people from the Sahara belt, influenced by Middle Eastern agriculturalists, moved south and interacted with yam “protoculturalists” and from this relationship yam-based agriculture developed;
5. and finally, around 2,500 B.P., with the advent of ironworking, West Africa people could expand deeper into the forest which ecologically favored yam over grain crops, and yam growing populations could achieve numerical superiority over grain farmers and create complex culture systems.

Although this model seems reasonable for the most part, the problems of documenting the domestication of West African yams are similar, and in some cases identical, to those associated with Southeast Asia. Here, too, primary evidence is lacking (Shaw 1976: 108-53).

Preliminary research on yam ennoblement, which was begun in 1977 in Nigeria, has led to the discovery that digging wild yams even with modern tools like machetes, shovels, and spades, let alone digging sticks, was arduous work (Chikwendu and Okezie 1989:

345). Wild yams could not be excavated like domesticated ones. They have long, sharp thorns all over their roots, and in addition to cutting through the yam roots, one has to cut through the tangled roots of the forest itself. A pick-like tool would only get caught between the roots. Trenching around a yam patch was the best procedure, but it still took several days just to dig up the first yam (Chikwendu and Okezie 1989: 345). This finding in turn casts some doubt on Coursey's proposal (1967: 13, 1975: 203) that the pick-like stone tools and Lupemban “hoes” of the Sangoan period were used for grubbing yams.

As with research on the Southeast Asian yams, indirect evidence like forest clearance and linguistics is our main avenue of inference. M. A. Sowunmi (1985: 127-9) reports significant changes in pollen counts from a Niger Delta soil core occurring around 850 B.C. He notes a sudden increase in oil-palm pollen and an increase in weed pollens associated with cultivated land, accompanied by a decrease in pollen of rain forest components, such as *Celtis* sp. and *Myrianthus arboreus*. Because there is no evidence of environmental change at that time, he concludes that humans artificially opened the forest for agricultural purposes. Because oil palm and yams are the main cultivars of aboriginal West African agriculture, he believes that these data document their appearance on a large scale.

It should be noted that, on the one hand, iron hoes, axes, and knives appeared in Nigeria (with the Nok complex) only about 300 years later, around 550 B.C. On the other hand, the site of Iwo Eleru has polished groundstone axes, dating as early as 4000 B.C., that could have been used in forest clearance, and Coursey (1967: 197-205, 1976b: 397) argues that yams were grown before the development of an iron technology because many of the peoples of West Africa have strong prohibitions against the use of iron tools in their important New Yam festivals.

Linguistics, as mentioned, is another source of information. Kay Williamson's study (1970: 156-67) of names for food plants within the Kwa Branch of the Niger-Congo family (spoken in the Niger Delta region) isolated “three main layers of names; ancient West African plants, crops of the Malaysian complex introduced long ago, and more recent introductions over the last five hundred years” (Williamson 1970: 163). The oil palm and the yam (*D. cayenensis-rotundata* complex) belong to the oldest layer; the banana, plantain, and water yam (*D. alata*) occurred in the Malaysian layer; and such plants as maize and manioc (cassava) are more recent introductions from the New World some five hundred years ago.

Williamson does not assess the antiquity of the words for yam and oil palm in calendar years, but P. J. Darling (1984: 65) proposes that the Proto-Kwa language dates from between 4,000 and 10,000 years ago. Although he calls these Proto-Kwa speakers Late-Stone-Age hunter-gatherers, it seems clear that as

they had words for major domesticated plants, they must already have been farmers. It is interesting that the more recent end of this date range matches Coursey's model for yam "protoculturalists" quite well. Finally, Proto-Niger-Congo not only has the word for yam (and cow and goat) but also the root meaning "to cultivate," and Proto-Niger-Congo may date back to at least 6000 B.C. (Ehret 1984: 29–30).

Thus, the evidence, though indirect, does point to the existence of yam usage and the concept of cultivation at around 6000 B.C. and forest clearance at about 850 B.C., presumably for the purpose of producing oil palms and yams on a wider scale. All of this in turn suggests an antiquity for agriculture in Africa far greater than believed by many scholars, which probably can best be explained in terms of an independent agricultural development in West Africa. Yet the West African yams had no dispersal beyond their region of origin until they were transferred to the tropical New World in association with the slave trade.

Summary

Three main waves of dispersal are associated with the spread of the sweet potato, in what Yen (1982: 20–1) calls the *kumara*, *kamote*, and *batatas* lines of transfer. The best-known and documented transfer was the post-Columbian spread via Europeans associated with the latter two lines. The Spanish, or *kamote* line, introduced the sweet potato into Europe, the Philippines, Guam, and Malaysia. From the Philippines it was then carried to China and from China ultimately to Japan. English immigrants transmitted it to the United States, English traders brought it to Japan (though it was not accepted), and English missionaries introduced it in parts of Melanesia and Australian New Guinea.

The Portuguese, or *batatas* line, introduced the sweet potato into India and Africa, Ambon, Timor, the northern Moluccas, and Cebu. The African introduction was from the Portuguese into Angola and Mozambique, as well as to Africa via Bombay through English associations with that trade center in India. Apparently the plant was carried from Burma to China after the Indian introduction.

The *kumara* line, the earliest, is associated with the appearance of the sweet potato in Oceania. This transfer has intrigued scholars for years. New primary evidence, combined with linguistic and historical data, point to a pre-Columbian spread somewhere into eastern Polynesia or even into northern Melanesia by the time of Christ. From this region the plant spread to all points of the Polynesia triangle. It then moved to parts of Melanesia via the Polynesians, and traveled from Melanesia into New Guinea. The transfer into New Guinea was probably accomplished by Melanesians, possibly bird of paradise hunters or migrants settling on the southeast coast. Though some specific areas of New Guinea received the plant from

Europeans, in general it was first spread by Melanesians and then by Papuans from the coast into the highlands, probably through the Markham Valley. The way in which early sweet potatoes reached New Guinea cannot presently be determined, but in the light of the Yap data it could be earlier than generally supposed.

The establishment of the sweet potato in many areas of Micronesia, parts of central Polynesia, and sections of Dutch New Guinea, including Lakor and the Arru and Tenimber islands, was prevented by ecological conditions unsuitable to its growth.

Yams also had several waves of dispersal. The Southeast Asian yams moved through the region beginning about 4500 B.C., and on into Oceania by 1500 B.C. They arrived in India and China in the first millennium A.D., and early in the second millennium entered Madagascar. From East Africa they moved to West Africa by the sixteenth century, and at the end of the sixteenth century came to the tropical New World.

General Conclusions

This survey on the problem of the origin and dispersal of the sweet potato and of yams indicates the following. First, the sweet potato originated in northwestern South America around 8000 B.C., in association with the initial development of tropical-forest root crop agriculture. The actual botanical ancestor is probably the result of various *n* and *2n* crosses within the *I. trifida* complex.

Primary evidence of the pre-Magellan introduction of the sweet potato into central Polynesia is established at around A.D. 1000, and even earlier, A.D. 50, in Micronesia on Yap. When combined with the archaeology of Oceania, these data suggest, conservatively, that the plant arrived in eastern Polynesia, maybe in the Fiji area, by about 500 B.C. Alternatively, the plant was dispersed by Lapita people sometime between 1500 to 500 B.C. during their movement through Melanesia. From Melanesia it was carried to New Guinea by Melanesians at an unknown date, but this could have taken place prior to the arrival of the Europeans.

The transference between Polynesia and the New World would seem to have been the result of either human accident or natural causes. An introduction by the casting up of a vessel upon some island of eastern Polynesia is possible, but it is equally possible that the plant was spread by natural agents, such as birds carrying seeds or by floating seed capsules. Both these hypotheses need further examination.

The post-European introduction of the sweet potato into Africa, North America, Europe, India, China, Japan, the Philippines, the Moluccas, and other islands in the Indonesian area was the result of Spanish, Portuguese, and English trade, exploration, colonization, and missionization.

The five ennobled yams were domesticated in Southeast Asia, West Africa, and tropical America, respectively, although the last region is not especially important to this study. Southeast Asian yams were probably domesticated before 4500 B.C., whereas the West African yams could be as old as 6000 B.C. but were probably domesticated by the first millennium B.C. The possible botanical ancestors of these yams are a subject of debate, and considerable cytological and taxonomic research is needed before this issue will be resolved. Needless to say, these ancestors will be found to have been native to each respective area.

Patricia J. O'Brien

I wish to thank Dr. Roger Green of the University of Auckland, Dr. Helen Leach of the University of Otago, Dr. Patrick V. Kirch of the University of California at Berkeley, and Dr. Donald Ugent of Southern Illinois University for kindly answering my questions about their research, and also for generously sharing with me reprints of their work.

Bibliography

- Austin, Daniel F. 1978. The *Ipomoea batatas* complex - I. Taxonomy. *Bulletin of the Torrey Botanical Club* 105: 114-29.
1983. Variability in sweet potatoes in America. In *Breeding new sweet potatoes in the tropics*, ed. F. W. Martin, 15-25. Mayaguez, P.R.
- Ayensu, Edward S., and D. G. Coursey. 1972. Guinea yams, the botany, ethnobotany, use and possible future of yams in West Africa. *Economic Botany* 26: 301-18.
- Bellwood, Peter. 1978. *The Polynesians: Prehistory of an island people*. London.
1985. *Prehistory of the Indo-Malaysian archipelago*. Sydney.
1991. The Austronesian dispersal and the origin of languages. *Scientific American* 265: 88-93.
- Biggs, Bruce G. 1972. Implications of linguistic subgrouping with special reference to Polynesia. In *Studies in oceanic culture history*, Vol. 3, ed. R. C. Green and M. Kelly, 143-52. Honolulu.
- Blust, Robert. 1976. Austronesian culture history: Some linguistic inferences and their relations to the archaeological record. *World Archaeology* 8: 19-43.
- Boxer, C. R. 1969. *Four centuries of Portuguese expansion, 1415-1825: A succinct survey*. Berkeley, Calif.
- Braidwood, Robert J. 1960. The agricultural revolution. *Scientific American* 203: 130-48.
- Brand, Donald D. 1971. The sweet potato: An exercise in methodology. In *Man across the sea*, ed. C. L. Riley, J. C. Kelley, C. W. Pennington, and R. L. Rands, 343-65. Austin, Tex.
- Broertjes, C., and A. M. van Harten. 1978. *Application of mutation breeding methods in the improvement of vegetatively propagated crops*. Amsterdam.
- Bulmer, Ralph. 1966. Birds as possible agents in the propagation of the sweet potato. *The Emu* 65: 165-82.
- Bulmer, Susan. 1989. Gardens in the south: Diversity and change in prehistoric Maori agriculture. In *Foraging and farming: The evolution of plant exploitation*, ed. D. R. Harris and G. C. Hillman, 688-705. London.
- Candolle, Alphonse de. 1959. *The origin of cultivated plants*. New York.
- Carter, George F. 1977. A hypothesis suggesting a single origin of agriculture. In *Origins of agriculture*, ed. C. A. Reed, 89-133. The Hague.
- Chang, K. C. 1981. In search of China's beginnings: New light on an old civilization. *American Scientist* 69: 148-60.
- Chikwendu, V. E., and C. E. A. Okezie. 1989. Factors responsible for the ennoblement of African yams: Inferences from experiments in yam domestication. In *Foraging and farming*, ed. D. R. Harris and G. C. Hillman, 418-25. London.
- Clark, Ross. 1979. Language. In *The prehistory of Polynesia*, ed. J. D. Jennings, 249-70. Cambridge, Mass.
- Colenso, W. 1880. On the vegetable food of the ancient New Zealanders. *Transactions of the New Zealand Institute* 13: 3-38.
- Conklin, Harold C. 1963. The Oceanian-African hypotheses and the sweet potato. In *Plants and the migrations of Pacific peoples*, ed. J. Barrau, 129-36. Honolulu.
- Cooley, J. S. 1951. The sweet potato - its origin and primitive storage practices. *Economic Botany* 5: 378-86.
- Coursey, D. G. 1967. *Yams*. London.
1975. The origins and domestication of yams in Africa. In *Gastronomy: The anthropology of food and food habits*, ed. M. L. Arnott, 187-212. The Hague.
- 1976a. Yams. In *Evolution of crop plants*, ed. N. W. Simmonds, 70-4. London.
- 1976b. The origins and domestication of yams in Africa. In *Origins of African plant domestication*, ed. J. R. Harlan, J. M. J. de Wet, and A. B. Stemler, 383-408. The Hague.
1980. The origins and domestication of yams in Africa. In *West African culture dynamics*, ed. B. K. Swartz, Jr., and R. E. Dumett, 67-90. The Hague.
- Damm, Hans. 1961. Die Süßkartoffel (Batate) im Leben der Völker Neuguineas. *Zeitschrift für Ethnologie* 84: 208-23.
- Darling, P. J. 1984. *Archaeology and history in southern Nigeria*, Part 1. Oxford.
- Dillehay, Tom. 1984. A late ice-age settlement in southern Chile. *Scientific American* 251: 106-17.
- Dillehay, Tom, and Michael B. Collins. 1988. Early cultural evidence from Monte Verde in Chile. *Nature* 332: 150-2.
- Dixon, Roland B. 1932. The problem of the sweet potato in Polynesia. *American Anthropologist* 34: 40-66.
- Early Polynesian migration to New Zealand reported. 1966. *New York Times*, May 12.
- Ehret, C. 1984. Historical/linguistic evidence for early African food production. In *From hunters to farmers*, ed. J. D. Clark and S. A. Brandt, 26-39. Berkeley, Calif.
- Engel, Frederic. 1970. Exploration of the Chilca canyon, Peru. *Current Anthropology* 11: 55-8.
- Frazer, Roger M. 1964. Changing Fiji agriculture. *The Australian Geographer* 9: 148-55.
- Golson, Jack. 1959. Culture change in prehistoric New Zealand. In *Anthropology in the South Seas*, ed. J. D. Freeman and W. R. Geddes, 29-74. New Plymouth, New Zealand.
1977. No room at the top: Agricultural intensification in the New Guinea highlands. In *Sunda and Sabul*, ed. J. Allen, J. Golson, and R. Jones, 601-38. London.
1989. The origins and development of New Guinea agriculture. In *Foraging and farming*, ed. D. R. Harris and G. C. Hillman, 678-87. London.
- Gorman, C. F. 1969. Hoabinhian: A pebble-tool complex with early plant associations in southeast Asia. *Science* 163: 671-73.

1970. Excavations at Spirit cave, north Thailand: Some interpretations. *Asian Perspectives* 13: 79-108.
- Green, Roger. 1979. Lapita. In *The prehistory of Polynesia*, ed. J. D. Jennings, 27-60. Cambridge, Mass.
- Groube, Les. 1989. The taming of the rain forests: A model for late Pleistocene forest exploitation in New Guinea. In *Foraging and farming*, ed. D. R. Harris and G. C. Hillman, 292-304. London.
- Hallam, Sylvia J. 1989. Plant usage and management in south-west Australian Aboriginal societies. In *Foraging and farming*, ed. D. R. Harris and G. C. Hillman, 136-51. London.
- Handy, E. S. C. 1940. *The Hawaiian planter*, Vol. 1. Honolulu.
- Harlan, Jack R. 1977. The origins of cereal agriculture in the old world. In *Origins of agriculture*, ed. C. A. Reed, 357-83. The Hague.
- Harris, David R. 1972. The origins of agriculture in the tropics. *American Scientist* 60: 180-93.
- Hather, Jon, and Patrick V. Kirch. 1991. Prehistoric sweet potato [*Ipomoea batatas*] from Mangaia Island, central Polynesia. *Antiquity* 65: 887-93.
- Hawkes, J. G. 1989. The domestication of roots and tubers in the American tropics. In *Foraging and farming*, ed. D. R. Harris and G. C. Hillman, 481-503. London.
- Hedrick, U. P. 1919. *Sturtevant's notes on edible plants*. Albany, N.Y.
- Hill, H. Edward, and John Evans. 1989. Crops of the Pacific: New evidence from the chemical analysis of organic residues in pottery. In *Foraging and Farming*, ed. D. R. Harris and G. C. Hillman, 418-25. London.
- Ho, Ping-Ti. 1955. The introduction of American food plants into China. *American Anthropologist* 57: 191-201.
- Ho, Ping-Ti. 1977. The indigenous origins of Chinese agriculture. In *Origins of agriculture*, ed. C. A. Reed, 413-84. The Hague.
- Hollyman, K. J. 1959. Polynesian influence in New Caledonia: The linguistic aspects. *Journal of the Polynesian Society* 68: 357-89.
- Hornell, James. 1946. How did the sweet potato reach Oceania? *The Journal of the Linnean Society* 53: 41-62.
- Hunt, Terry L. 1981. New evidence for early horticulture in Fiji. *Journal of the Polynesian Society* 90: 259-66.
- Jennings, Jesse D. 1979. Introduction. In *The prehistory of Polynesia*, ed. J. D. Jennings, 1-5. Cambridge, Mass.
- Kobayashi, Masashi. 1984. The *Ipomoea trifida* complex closely related to the sweet potato. In *Symposium of the International Society of Tropical Root Crops*, ed. F. S. Shideler and H. Rincon, 561-9. Lima.
- Kolff, D. H. 1840. *Voyages of the Dutch brig-of-war Dourga*, trans. G. W. Earl. London.
- Lathrap, Donald W. 1962. Yarinacocha: Stratigraphic excavations in the Peruvian Montana. Ph.D. dissertation, Harvard University.
1970. *The Upper Amazon*. New York.
1977. Our father the cayman, our mother the gourd: Spinden revisited, or a unitary model for the emergence of agriculture. In *Origins of agriculture*, ed. C. A. Reed, 713-52. The Hague.
- Law, R. Garry. 1969. Pits and kumara agriculture in the South Island. *Journal of the Polynesian Society* 78: 223-51.
- Lawton, H. W., P. J. Wilke, M. DeDecker, and W. M. Mason. 1976. Agriculture among the Paiute of Owens valley. *The Journal of California Anthropology* 3: 13-50.
- Leach, Helen M. 1979. The significance of early horticulture in Palliser Bay for New Zealand prehistory. In *Prehistoric man in Palliser Bay*, ed. B. F. Leach and H. M. Leach, 241-48. Wellington.
1984. *1000 years of gardening in New Zealand*. Wellington.
1987. The land, the provider: Gathering and gardening. In *From the beginning*, ed. J. Wilson, 85-94. Wellington.
- MacNeish, Richard S. 1964. Ancient Mesoamerican civilization. *Science* 143: 531-537.
1977. The beginnings of agriculture in central Peru. In *Origins of agriculture*, ed. C. A. Reed, 753-801. The Hague.
- Marchant, Alexander. 1941. Colonial Brazil as a way station for Portuguese India fleets. *The Geographical Review* 31: 454-65.
- Martin, Franklin W., and Alfred Jones. 1986. Breeding sweet potatoes. *Plant Breeding Reviews* 4: 313-45.
- Meacham, William. 1988. On the improbability of Austronesian origins in South China. *Asian Perspectives* 26: 89-106.
- Mendana, A. de. 1901. *The discovery of the Solomon Islands*. 2 vols. London.
- Merrill, Elmer D. 1954. The botany of Cook's voyages. *Chronica Botanica* 14: 161-384.
- Miege, J. 1982. Appendix note on the *Dioscorea cayenensis* Lamk. and *D. rotundata* Poir. species. In *Yams/Ignames*, ed. J. Miege and S. N. Lyonga, 377-83. Oxford.
- Nishiyama, Ichizo. 1961. The origin of the sweet potato plant. In *Abstracts of Symposium Papers: Tenth Pacific Science Congress*, 137-8. Honolulu.
1963. The origin of the sweet potato plant. In *Plants and the migrations of Pacific peoples*, ed. J. Barrau, 119-28. Honolulu.
- Nishiyama, I., T. Miyazakim, and S. Sakamoto. 1975. Evolutionary autopoloidy in the sweet potato (*Ipomoea batatas* [L.] Lam.) and its progenitors. *Euphytica* 24: 197-208.
- O'Brien, Patricia J. 1972. The sweet potato: Its origin and dispersal. *American Anthropologist* 74: 342-65.
- Orjeda, G., R. Freyre, and M. Iwanaga. 1990. Production of 2n pollen in diploid *Ipomoea trifida*, a putative wild ancestor of sweet potato. *Journal of Heredity* 81: 462-7.
- Patterson, Thomas C., and Edward P. Lanning. 1964. Changing settlement patterns on the central Peruvian coast. *Nawpa Pacha* 2: 113-23.
- Patterson, Thomas C., and M. Edward Moseley. 1968. Late preceramic and early ceramic cultures of the central coast of Peru. *Nawpa Pacha* 6: 115-33.
- Pawley, Andrew, and Kaye Green. 1971. Lexical evidence for the proto-Polynesian homeland. *Te Reo* 14: 1-35.
- Pickersgill, Barbara, and Charles B. Heiser, Jr. 1977. Origins and distribution of plants domesticated in the New World tropics. In *Origins of agriculture*, ed. C. A. Reed, 803-35. The Hague.
- Polynesian settlement of New Zealand. 1966. *New Zealand News* 21: 1-8.
- Price, R. H. 1896. *Sweet potato culture for profit*. Dallas, Texas.
- Purseglove, J. W. 1965. The spread of tropical crops. In *The genetics of colonizing species*, ed. H. G. Baker and G. L. Stebbins, 375-86. New York.
- Reynolds, Robert G. 1986. An adaptive computer model for the evolution of plant collecting and early agriculture in the eastern valley of Oaxaca. In *Guila Naquitz*, ed. K. V. Flannery, 263-89. Orlando, Fla.
- Roosevelt, A. C., R. A. Housley, M. Imazio da Silveira, et al. 1991. Eighth millennium pottery from a prehistoric shell midden in the Brazilian Amazon. *Science* 254: 1621-4.
- Rosendahl, P., and D. E. Yen. 1971. Fossil sweet potato

- remains from Hawaii. *Journal of the Polynesian Society* 80: 379-85.
- Safford, William E. 1925. The potato of romance and of reality. *Journal of Heredity* 16: 217-29.
- Saggers, Sherry, and Dennis Gray. 1984. The 'Neolithic problem' reconsidered: Human-plant relationships in northern Australia and New Guinea. *Asian Perspectives* 25: 99-125.
- Sauer, Carl O. 1952. *Agricultural origins and dispersals*. New York.
- Schmitz, Carl A. 1960. *Historische Probleme in nordost New Guinea*. Wiesbaden.
- Schoenwetter, James. 1990. Lessons from an alternative view. In *Powers of observation: Alternative views in archaeology*, ed. S. M. Nelson and A. B. Kehoe, 103-12. Washington, D.C.
- Schoenwetter, James, and Landon Douglas Smith. 1986. Pollen analysis of the Oaxaca archaic. In *Guila Naquitz*, ed. K. V. Flannery, 179-238. Orlando, Fla.
- Serpenti, L. M. 1965. *Cultivators in the swamp: Social structure and horticulture in a New Guinea society*. Assen, Netherlands.
- Sharma, A. K., and Archana Sharma. 1957. Vegetatively reproducing plants - their means of speciation. *Science and Culture* 22: 628-30.
- Shaw, Thurston. 1976. Early crops in Africa: A review of the evidence. In *Origins of African plant domestication*, ed. J. R. Harlan, J. M. J. de Wet, and A. B. L. Stemler, 108-53. The Hague.
1978. *Nigeria: Its archaeology and early history*. London.
- Simon, Edmund. 1914. The introduction of the sweet potato into the Far East. *Transactions on the Asiatic Society of Japan* 42: 711-24.
- Skjolsvold, Arne. 1961. Site E-2, a circular dwelling, Anakena. In *Archaeology of Easter Island*, Vol. 1., ed. T. Heyerdahl and E. N. Ferdon, Jr., 295-303. Santa Fe, N.Mex.
- Solheim, Wilhelm G., II. 1969. Mekong valley flooding. In *Smithsonian Institution Center for Short-Lived Phenomena, "Report on Mekong Valley Flooding"* (no. 58-69). Washington, D.C.
1988. The Nusantara hypothesis: The origin and spread of Austronesian speakers. *Asian Perspectives* 26: 77-88.
- Sowunmi, M. A. 1985. The beginnings of agriculture in West Africa: Botanical evidence. *Current Anthropology* 26: 127-9.
- Steward, Julian H. 1930. Irrigation without agriculture. *Papers of the Michigan Academy of Sciences, Arts and Letters* 12: 149-56.
- Sutton, Douglas 1984. The Pouerua project: Phase II, an interim report. *New Zealand Archaeological Association Newsletter* 27: 30-8.
- Turner, C. G. 1989. Teeth and prehistory in Asia. *Scientific American* 260: 88-96.
- Ugent, Donald, Tom Dillehay, and Carlos Ramirez. 1987. Potato remains from a late Pleistocene settlement in southcentral Chile. *Economic Botany* 41: 17-27.
- Ugent, Donald, and Linda W. Peterson. 1988. Archaeological remains of potato and sweet potato in Peru. *International Potato Center Circular* 16: 1-10.
- Ugent, Donald, Sheila Pozorski, and Thomas Pozorski. 1981. Prehistoric remains of the sweet potato from the Casma valley of Peru. *Phytologia* 49: 401-15.
1982. Archaeological potato tuber remains from the Casma valley of Peru. *Economic Botany* 36: 182-92.
- Watson, James B. 1965. The significance of a recent ecological change in the central highlands of New Guinea. *Journal of the Polynesian Society* 74: 438-50.
- White, J. Peter, and James F. O'Connell. 1982. *The prehistory of Australia, New Guinea and Sabul*. Sydney.
- Williamson, Kay. 1970. Some food plant names in the Niger delta. *International Journal of American Linguistics* 36: 156-67.
- Yen, Douglas E. 1960. The sweet potato in the Pacific: The propagation of the plant in relation to its distribution. *Journal of the Polynesian Society* 69: 368-75.
1961. The adaption of kumara by the New Zealand Maori. *Journal of the Polynesian Society* 70: 338-48.
1973. Ethnobotany from the voyages of Mendana and Quiros in the Pacific. *World Archaeology* 5: 32-43.
1974. *The sweet potato and Oceania: An essay in ethnobotany*. Honolulu.
1976. Sweet potato. In *Evolution of crop plants*, ed. N. W. Simmonds, 42-5. London.
1977. Hoabinhian horticulture? The evidence and the questions from northeast Thailand. In *Sunda and Sabul*, ed. J. Allen, J. Golson, and R. Jones, 567-99. London.
1982. Sweet potato in historical perspective. In *Sweet potato, proceedings of the first international symposium*, ed. R. L. Villareal and T. D. Griggs, 17-30. Shanhua, Taiwan.
- Zavala, Silvio. 1964. New world contacts with Asia. *Asian Studies* 2: 213-22.

II.B.6 🌿 Taro

Taro is the common name of four different root crops that are widely consumed in tropical areas around the world. Taro is especially valued for its starch granules, which are easily digested through the bloodstream, thus making it an ideal food for babies, elderly persons, and those with digestive problems. It is grown by vegetative propagation (asexual reproduction), so its spread around the world has been due to human intervention. But its production is restricted to the humid tropics, and its availability is restricted by its susceptibility to damage in transport.

Taro is most widely consumed in societies throughout the Pacific, where it has been a staple for probably 3,000 to 4,000 years. But it is also used extensively in India, Thailand, the Philippines, and Southeast Asia, as well as in the Caribbean and parts of tropical West Africa and Madagascar (see Murdock 1960; Petterson 1977). Moreover, in the last quarter of the twentieth century taro entered metropolitan areas such as Auckland, Wellington, Sydney, and Los Angeles, where it is purchased by migrants from Samoa and other Pacific Island nations who desire to maintain access to their traditional foods (Pollock 1992).

Although taro is the generic Austronesian term for four different roots, true taro is known botanically as *Colocasia esculenta*, or *Colocasia antiquorum* in



some of the older literature. We will refer to it here as *Colocasia taro*. False taro, or giant taro, is the name applied to the plant known botanically as *Alocasia macrorrhiza*. It is less widely used unless other root staples are in short supply. We will refer to it as *Alocasia taro*.

Giant swamp taro is the name used for the plant known as *Cyrtosperma chamissonis*. This is a staple crop on some atolls, such as Kiribati, and is also grown in low-lying areas of larger islands. We will refer to it as *Cyrtosperma taro*. The fourth form of taro has been introduced into the Pacific and elsewhere much more recently. It is commonly known as tannia, kongkong, or American taro, but its botanical name is *Xanthosoma sagittifolium*. It has become widely adopted because it thrives in poorer soils and yields an acceptable food supply. We will refer to it as *Xanthosoma taro*.

The starch-bearing root, termed corm or cormlet, of these four plants makes a major contribution to the human diet, and even the leaves are occasionally used as food, as in the case of young *Colocasia* leaves. The leaves are also employed as coverings for earth ovens and as wrappings for puddings that are made from the grated starch root and baked in an earth oven. The large leaves of *Cyrtosperma taro* also provide a good substitute for an umbrella.

Botanical Features

All the taros are aroids of the Araceae family, so we would expect to find close similarities among them. They grow in tropical climates with an adequate year-round rainfall. All taros must be propagated vegetatively, as they do not have viable seeds. Consequently their production for food crops has been engineered by human intervention. Both the corms and the leaves are acrid to some degree, particularly before they are cooked, and cause an unpleasant irritation of the skin and mouth (Tang and Sakai 1983; Bradbury and Holloway 1988).

Colocasia Taro

The *Colocasia taro* plant consists of an enlarged root or corm, a number of leafstalks, and leaves. The leaves are the main visible feature distinguishing *Colocasia* from the other taros, particularly *Xanthosoma* (see Figure II.B.6.1). The *Colocasia* leaf is peltate, or shield-shaped, with the leafstalk joining the leaf about two-thirds of the way across it. Varieties of *Colocasia taro* differ in the color of the leafstalk, the shape and color of the leaf and veins, and the number of fully developed leaves. They also differ in the shape, flesh color, and culinary qualities of their tubers. The varieties are recognized by individual names, particularly in those societies in the Pacific where the plant is a major foodstuff. For example, 70 local names were recorded in Hawaii and 67 in Samoa (Massal and Barrau 1956; Lambert 1982). Indeed, fully 722 accessions have been recorded in collections of root crops in South Pacific countries (Bradbury and Holloway 1988).

Colocasia taro can be grown on flooded or irrigated land and on dry land. The planting material consists of the corm plus its leafstalks, minus the leaves. Taros in Fiji are sold this way in the market, so that the purchaser can cut off the root for food and plant the topmost part of the root plus leafstalks, known as a sett, to grow the next crop of taro. Harvesting one taro root therefore yields the planting material for the next crop. The root takes about 7 to 10 months to mature. Once the corm has been cut or damaged, taro rots quickly, making it difficult to ship to distant markets.

Alocasia Taro

Sometimes known as giant taro, or *kape* in Polynesian languages, *Alocasia taro* is a large-leaved plant that is grown for its edible stem rather than for the root. The fleshy leaves are spear-shaped and can reach more than a meter in length. The stem and central vein of the leaf form a continuous line. The leaves of this taro are not usually eaten.

The edible part is the large and long (a half meter or more), thickened underground stem that may weigh 20 kilograms (kg). It is peeled and cut into pieces to be cooked in an earth oven or boiled.

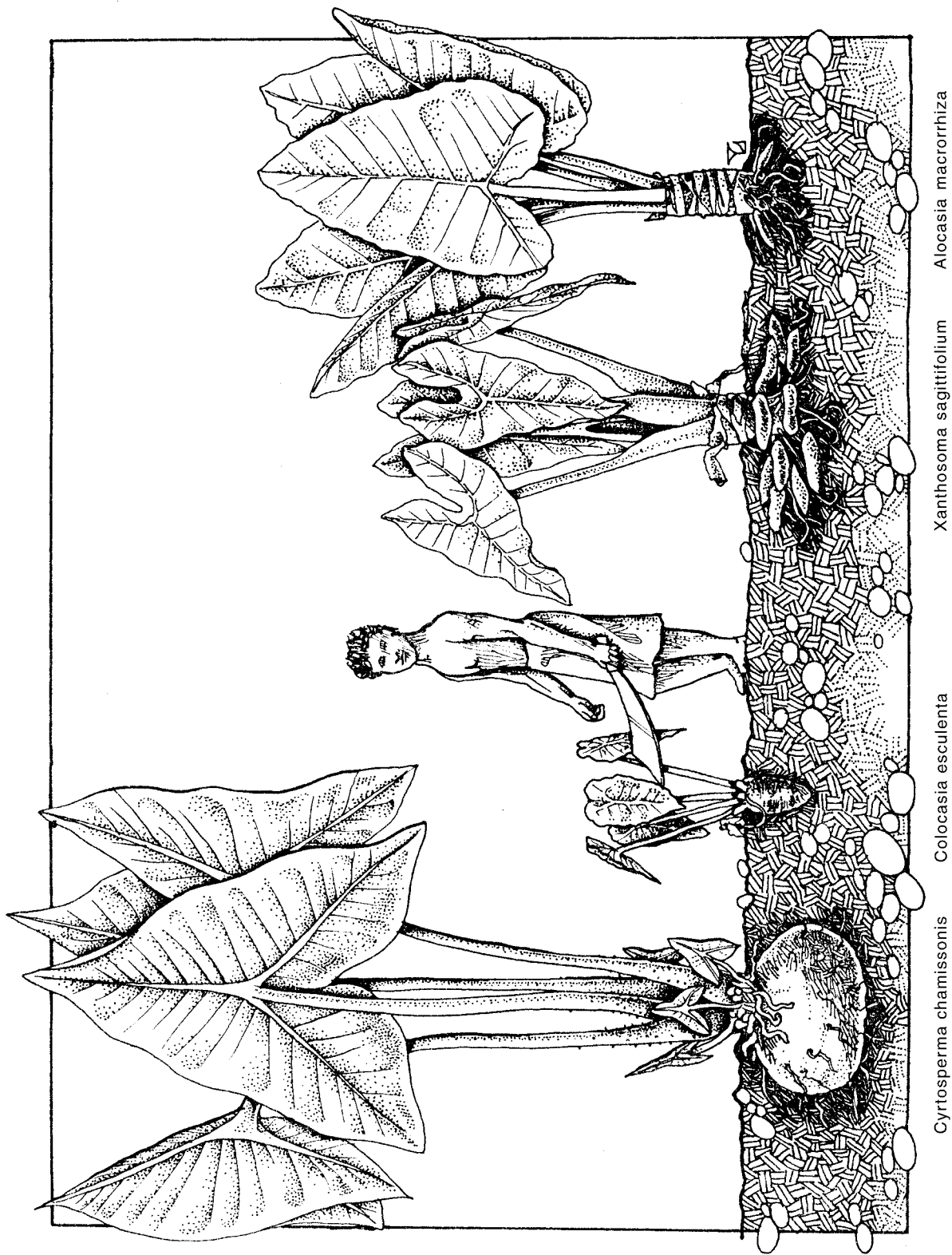


Figure II.B.6.1. Four types of taros. (Illustration by Tim Galloway).

Alocasia taros are very acrid, as they contain a high concentration of calcium oxalate crystals in the outer layers of the stem. These crystals are set free by chewing and can cause irritation in the mouth and throat if the stem is not thoroughly cooked. The calcium oxalate content increases if the plant is left in the ground too long. For this reason some societies consider *Alocasia taro* fit for consumption only in emergencies. But in Tonga, Wallis, and Papua New Guinea, varieties with very low oxalate content have been selectively grown to overcome this problem (Holo and Taumoeofalau 1982). Twenty-two accessions are held in collections of root crops for the South Pacific (Bradbury and Holloway 1988).

Cyrtosperma Taro

Cyrtosperma taro, called giant swamp taro, also has very large leaves, sometimes reaching 3 meters in height. In fact, a swamp taro patch towers over those working in it. The leaves are spear-shaped and upright, the central vein forming a continuous line with the stem. The edible corm grows to some 5 kg in size, depending on the variety.

This taro prefers a swampy environment and will withstand a high level of water, provided it is not inundated by seawater. It is grown in Kiribati under cultivation techniques that have been carefully developed over several hundred years (Luomala 1974).

Cyrtosperma taro is a highly regarded foodstuff in Kiribati and in Yap, as well as on other atolls in Micronesia and in the Tuamotus. It is also cultivated in the Rewa district of southeast Fiji, and evidence of its former cultivation can be found in northern Fiji, Futuna, and Wallis, where it is employed today as an emergency crop. It is rarely cultivated outside the Pacific.

Xanthosoma Taro

Xanthosoma taro, by contrast, is much more widespread than *Cyrtosperma taro* and may be found in many tropical areas, including those on the American and African continents, as well as in the Pacific, where several varieties are now being cultivated (Weightman and Moros 1982). Although it is a very recent introduction to the islands relative to the other three taros, it has become widely accepted as a household crop because it is easy to grow; it can be intercropped with other subsistence foods in a shifting cultivation plot, and it tolerates the shade of a partially cleared forest or of a coconut or pawpaw plantation. It cannot stand waterlogging.

The principal tuber of *Xanthosoma* is seldom harvested because it also contains calcium oxalate crystals. Rather, the small cormlets are dug up, some being ready 7 to 10 months after planting. These are about the size of a large potato, weighing up to half a kilo.

In appearance *Xanthosoma taro* is often confused

with *Colocasia taro*, the term "eddoe" being used for both. The main difference between them is in the leaf structure; the *Xanthosoma* leaf is hastate, an arrow or spearhead shape, and not peltate, so that the leafstalk joins the leaf at the edge, giving it a slightly more erect appearance than the *Colocasia* leaf (see Figure II.B.6.1). The distinctive feature of the *Xanthosoma* leaf is its marked vein structure together with a marginal vein.

Production

As already noted, all of the taros must be propagated vegetatively as none of them produce viable seeds naturally. They thus require human intervention both for introduction to a new area and for repeated production of a food supply. This factor further strengthens the likelihood of selection of particular varieties that have more desirable attributes as food, such as reduced acidity and suitability to particular growing conditions. The planting material for all the taros is either a sucker or the sett (consisting of the base of the petioles and a 1-centimeter section from the top of the corm).

Dryland or upland cultivation is the most widespread form of cultivation of *Colocasia taro*, which grows best in a warm, moist environment. It can be grown between sea level and 1,800 meters (m) where daily average temperatures range between 18 and 27 degrees Celsius (C), with rainfall of about 250 centimeters (cm) annually. The setts or suckers are placed in a hole made with a digging stick, with a recommended spacing between plants of 45 cm by 60 cm to produce good-sized corms. Yield will increase with higher planting density (De La Pena 1983).

Irrigated or wetland *Colocasia taro* is grown in prepared beds in which, to prevent weed seed germination, water is maintained at a level of 5 cm before planting and during the growth of the setts until the first leaves unfurl. The beds may be a few feet across, beside a stream or well-watered area, or they may be in an area some 2 to 3 acres in size, depending on the land and water available. After the first leaves appear, the beds are frequently covered with whole coconut fronds to shade the young plants from the sun.

Irrigated taro planted at a density of 100,000 plants/ha yields 123.9 tons/ha. With 10,000 plants/ha the yield is 41.4 tons/ha (De La Pena 1983: 169-75). Clearly, the more intense techniques developed for wetland cultivation produce a higher yield. But these techniques are suited only to areas where the right conditions pertain, and on the Pacific Islands such areas are limited because of the nature of the terrain. Dryland taro is, thus, more versatile.

Colocasia taro, whether upland or irrigated, may be harvested after a growing period ranging from 9 to 18 months, depending on the variety and the growing

conditions. In Fiji some varieties are harvestable at 9 to 11 months after planting, whereas in Hawaii harvest takes place from 12 to 18 months after planting in the commercial fields.

The subsistence farmer, by contrast, harvests only those taros needed for immediate household use. The farmer will cut off the setts and plant them in a newly cleared area of land or in a different part of the irrigated plot. Thus, any one household will have several taro plots at different stages of growth to maintain a year-round supply and to meet communal obligations for feasts and funerals.

Pests and pathogens are a greater problem for *Colocasia* taro than for the other varieties. Both the leaves and the corm are subject to damage during the growing period from a range of pests and biotic agents (Mitchell and Maddison 1983). In the Pacific, leaf rot and corm rot have been the most serious diseases, spreading rapidly through whole plantations, particularly in Melanesia (Ooka 1983). In fact, these diseases were so severe in the early 1970s that some societies in the Solomons ceased taro consumption and switched to the sweet potato. West Samoa has recently lost its entire crop due to these diseases.

Alocasia is interplanted with yams and sweet potatoes in Tonga and Wallis where it is grown in shifting cultivation plots. The planting material is usually the larger suckers (although cormlets may also be used). These are placed in holes 10 to 25 cm deep, preferably between July and September. If planted with spacing of 1.5 m by 1.5 m *Alocasia* will yield 31 tons/ha, with an average kape root stem weighing 8 to 10 kg and reaching 1.5 m in length. The plant suffers few pests, is weeded when convenient, and is harvested a year after planting (Holo and Taumoeofolau 1982:84).

Swamp *Cyrtosperma* production has been culturally elaborated in some Pacific societies so that it is surrounded by myth and secrecy (see Luomala 1974 for Kiribati). The planting material may be a sett or a young sucker, and in Kiribati this is placed in a carefully prepared pit that may be several hundred years old, to which mulch has been constantly added. Each new plant is set in a hole with its upper roots at water level, surrounded by chopped leaves of particular plants chosen by the individual planter and topped with black humic sand. It is encased in a basket of woven palm fronds, to which more compost mixture is added as the plant grows. The larger cultivars are spaced at 90 cm by 90 cm; smaller ones are spaced more closely.

A pit is likely to consist of *Cyrtosperma* plants at various stages of growth. A corm may be harvested after 18 months – or it may be left for 15 years, by which time it is very fibrous and inedible but still brings prestige to the grower's family when presented at a feast. Yield is uneven due to different cultivation practices but may reach 7.5 to 10 tons/ha. An individual corm may weigh 10 to 12 kg (Vickers and Untaman 1982).

Xanthosoma taro is best grown on deep, well-drained, fertile soils. It tolerates shade and so can be interplanted with other crops such as coconuts, cocoa, coffee, bananas, and rubber, or with subsistence crops such as yams. The planting material is the cormlet or a sett, but the former grows more quickly than the latter. These can be planted at any time of year but grow best if planted just before onset of the rainy season. If they are spaced at 1 m by 1 m, the yield is about 20 tons/ha. The plant, which has few pests or diseases, produces a number of cormlets the size of large potatoes. These can be harvested after six months, but they bruise easily, reducing storage time (Weightman and Moros 1982).

Different planting techniques have been developed over time in order to provide a foodstuff that suits both the palate of those eating it and the local growing conditions. Most taro (*dryland Colocasia*, *Alocasia*, and *Xanthosoma*) is grown under dryland conditions with reasonable rainfall, and it seems likely that the aroids were all originally dryland plants (Barrau 1965). Because the techniques involved in the cultivation of wetland *Colocasia* and swamp *Cyrtosperma* taro are more arduous than those of dryland cultivation, one suspects that these plants were encouraged to adapt to wetland conditions, probably to meet specific food tastes.

Origins

Colocasia and *Alocasia* taro are among the oldest of the world's domesticated food plants. Both apparently have an Asian origin, possibly in India or Burma (Massal and Barrau 1956), but because they consist of vegetal material that has no hard parts, they leave almost no trace in the archaeological record. As a consequence, there has been much room for debate about the early development of the taros.

Some of the debate has centered on the question of whether root crop domestication preceded that of cereals, such as millet and rice, in the Southeast Asia region. Most authorities now agree that root crops came first (e.g., Chang 1977), although C. Gorman (1977) considered rice and taro as sister domesticates. In an overview of the evidence, M. Spriggs (1982) argued that root crops, including taro, were early staples in Southeast Asia, with rice becoming a staple much later.

The time depth is also problematic. It has been suggested that the early agricultural phase of slash-and-burn, dryland cultivation in Southeast Asia took place some 8,000 to 10,000 years ago, with a sequence of dominant cultigens proceeding from root crops to cereals (Hutterer 1983). Dryland taro may have been cultivated for some 7,000 years or more, with wetland (*Colocasia*) taro forming part of a second stage of development of Southeast Asian food crops (Bellwood 1980).

Prehistorians have given much more attention to

wetland taro irrigation techniques and those used in the production of paddy rice than they have to dryland practices. This is because wetland techniques are said to mark technological innovation and, thus, to be associated with a more complex level of social organization than that required in the production of dryland taro or rice (e.g., Spriggs 1982 for Vanuatu; Kirch 1985 for Hawaii). Yet this does not necessarily mean that foodstuffs produced by irrigation had greater importance in the diet than their dryland counterparts. We need to know the importance of such crops in the food consumption and exchange systems of those people who chose to develop a more complex mode of production. For a food to be considered a staple, the proportion of dietary content is the important aspect, as opposed to the techniques of production or the size of the production units.

Ease of cooking may constitute another reason that taro preceded rice. A hole lined with stones, in which a fire was lit, became an oven with limited tools. The whole taro root could be cooked thoroughly in such an oven, together with fish or pork or other edibles. To cook rice, by contrast, required some form of utensil in which to boil the rice to make it edible, either in the form of cakes, or as a soup.¹

But if taro, whether *Colocasia* or *Alocasia*, had advantages that suggest it was an earlier domesticated foodstuff than rice, its disadvantages lay in its bulk and post-harvest vulnerability. However, advantages outweighed disadvantages, so it is not surprising that these two forms of taro spread as widely as they did across Oceania and into Africa.

Despite its antiquity, however, the origin of *Alocasia* taro has not attracted as much attention as that of *Colocasia* taro; similarly, we know more about the origin of *Xanthosoma* taro than we do about *Cyrtosperma* taro. This is partly because the *Alocasia* and *Cyrtosperma* taros are not as widely used as the other two and partly because, even where they are used as foods (save for *Cyrtosperma* in Kiribati and Yap), they are not the main foodstuff.

Alocasia taro has its origins either in India (Plucknett 1976) or in Sri Lanka (Bradbury and Holloway 1988) but has been grown since prehistory throughout tropical southeast Asia, as well as in China and Japan (Pettersen 1977). *Cyrtosperma*, by contrast, was said to have been first domesticated either in the Indo-Malaya region (Barrau 1965) or in Indonesia or Papua New Guinea, where it has wild relatives (Bellwood 1980). Both *Alocasia* and *Cyrtosperma* taros are abundant in the Philippines, where a number of different varieties of each are known (Pettersen 1977). *Cyrtosperma* remains, discovered there by archaeologists, suggest that it was cultivated at least from A.D. 358 (Spriggs 1982), indicating that there was sufficient time to develop a range of plant types yielding less acidic starch foods.

In their reconstruction of the early forms of the Austronesian language, A. Pawley and R. Green (1974)

include words for *Alocasia*, *Colocasia*, and *Cyrtosperma* taros, indicating a time depth in the Pacific of some 3,000 to 4,000 years. Thus, all three plants have probably been domesticated and exchanged for over 5,000 years in tropical Southeast Asia.

Xanthosoma taro differs from the other three aroids in having its homeland in tropical America. Little is known about the Xantharoids before the twentieth century, as J. Pettersen (1977) points out in her thesis on the dissemination and use of the edible aroids. But she does offer us one possible reconstruction of the spread of what she calls American taro. It was "a most ancient domesticate of the Western hemisphere," said to have originated in the Caribbean lowlands along the northern coast of South America.

By the time Europeans reached the Americas, *Xanthosoma* taro had diffused south into northwest South America and north across the Antilles and Central America where several varieties were known.² Exactly where *Xanthosoma* was consumed at this time, and who consumed it, is unclear, although later on taro roots served as an important foodstuff for slaves on sugar plantations.

Geographic Spread

The four aroids have spread around the tropical areas of the world, with *Colocasia* and *Xanthosoma* more widely cultivated than *Alocasia* and *Cyrtosperma* taros. A range of varieties of each has been developed by human selectivity, and all four are extensively utilized by the island societies of Oceania.

Colocasia taro was carried from its South Asia homeland in both an easterly and a westerly direction, probably some 6,000 years ago (Bellwood 1980). Moving east it became established in Thailand, Malaysia, Indonesia, and the Philippines and from there was carried by canoe into Papua New Guinea, the Marianas, and henceforth into Micronesia and Polynesia (Pettersen 1977; Yen 1980; Hutterer 1983; Pollock 1992). The Malay name *tales* is the base of the common term taro as widely used today. In the four areas of Oceania the *Colocasia* root has gained a reputation as a highly valued foodstuff that also has prestige value (though not as high as *Dioscorea* yams). Today it is still cultivated as a food in Hawaii, the Marquesas, Tahiti, the Cooks, the Solomons, and Papua New Guinea. It remains a major foodstuff in Samoa, Tonga, Wallis, Futuna, Fiji, and Vanuatu.³

The easterly spread of *Colocasia* taro across the islands of the Pacific is today associated by prehistorians with the development of Lapita culture some 6,000 years ago. Lapita culture is a construct by prehistorians of a period in the settlement of the Pacific, with the spread of a particular form of pottery as its hallmark. Associated with this culture is the cultivation of *Colocasia* taro in particular, but *Alocasia* taro as well.

How sophisticated irrigation technology was introduced by these people moving out of Southeast Asia

has not been clearly established. It seems likely that dryland taro could have been introduced to a wider range of environments in the Pacific and, thus, was in existence earlier than irrigated or wetland taro.

For our purposes the important consideration is how these production techniques influenced the acceptability of the crop as a foodstuff. Because people in the Pacific today assert that irrigated taro is softer and less acrid, it is likely that the wetland Colocasia taro has undergone more specific selection than the dryland version, depending on how the root was prepared as a food. For the most part it was cooked whole in the earth oven, to be eaten in slices with accompaniments (as discussed in the next section). But in Hawaii, Rapa, and a few small islands it was employed mainly as *poi*, a fermented product made from either upland or wetland taro. The dietary uses of Colocasia taro are thus likely to have influenced which plants were selected for replanting and the techniques (whether wet or dry) of production. Moreover, appropriate cooking techniques had to be developed, as well as methods for overcoming the acidity.

In its westward spread, Colocasia taro was planted in India, where it still forms part of the diets of some societies today. It reached Madagascar where it became widely established between the first and eleventh centuries A.D. It was carried further westward in two branches, one along the Mediterranean and the other across Africa south of the Sahara.

The plant attained significance in West Africa where it has been grown ever since. In the Mediterranean region, it was flourishing in Egypt at the time of Alexander's expedition, where it was known as the Egyptian water lily and Egyptian beans. Virgil and Pliny both referred to the Colocasia plant in their writing, with the latter noting that "when boiled and chewed it breaks up into spidery threads" (quoted in Petterson 1977: 129). Colocasia taro has continued in importance in Egypt and also in Asia Minor and Cyprus until recently when it became too expensive to produce (Petterson 1977).

Colocasia taro reached the Iberian Peninsula probably by about A.D. 714, along with sugar cane. In both West Africa and Portugal the term *enyame* came to be applied to the Colocasia taro and was picked up by other Europeans as a generic name for all unfamiliar root crops. Thus, to English explorers yams meant any number of root crops, including both yams and Colocasia taro (Petterson 1977).

The latter also reached tropical America from the east, and it became a secondary foodstuff for Peruvian, Ecuadorean, and Amazonian peoples, as well as for those of the Caribbean, where it is known as dasheen or eddoe (Petterson 1977: 185). It seems likely that Colocasia taro was carried by the Iberians in their westward explorations and then brought from Africa to feed slaves during the Middle Passage.

By contrast, Alocasia taro has not traveled so widely. It was carried from Southeast Asia mainly into

the islands of Oceania where it has been domesticated on some atolls in Micronesia and Polynesia, as well as on some high islands, such as Samoa, Tonga, Wallis, and Fiji. On those islands where it is grown, a number of different varieties of the plant have been developed locally. In all these places the Alocasia taro makes a significant contribution to the diet.

Its westward spread from Indonesia to India was less prolific (Plucknett 1976). Today, Alocasia plants can be found growing as ornamentals in both tropical and subtropical areas, such as Florida (Petterson 1977) and in the northern part of the north island of New Zealand. They do not, however, serve as food.

Cyrtosperma is a small genus that is used as food almost exclusively in the Oceania region.⁴ But Xanthosoma taro spread from its homeland along the northern coast of South America out into the northern part of South America and through tropical Central America. It may have been of some importance as a food in Classic Maya civilization (c.A.D. 200-900).⁵

Spanish and Portuguese contacts with America led to the dispersal of Xanthoroids into Europe (they were grown in England as a curiosity in 1710) and, probably, into Africa where they may have died out and then been reintroduced. The root was allegedly brought to Sierra Leone in 1792 by former North American slaves who had fled to Nova Scotia after the American Revolution. However, the generally accepted date for the introduction of Xanthosoma taro to sub-Saharan Africa is April 17, 1843, when missionaries carried the American taro from the West Indies to Accra, Ghana.

It subsequently spread from West Africa to Uganda and through the Cameroons and Gabon to attain varying levels of importance. Beyond Africa, it traveled along Portuguese trading lines to India and the East Indies, following a similar route as the sweet potato. But it has not become an important food crop in Asia, except in the Philippines whence it spread to Malaysia (Petterson 1977).

Xanthosoma was introduced to the Pacific only in the last 200 years, probably via Hawaii (Barrau 1961) and Guam in contact with the Philippines (Pollock 1983). Its spread across the Pacific was aided by missionary activity as much as by island exchange, and the names used in Pacific societies today for Xanthosoma taro suggest the routes of transferal.

Taros As Food

Taro is a very important foodstuff in those societies that use it, both in the household and also for feasts and exchanges. But in terms of world food crops, the taros are considered of marginal importance. They rank behind bananas and root crops, such as cassava, sweet potatoes, and yams, in amounts consumed (Norman, Pearson, and Searle 1984: 221). Nonetheless, taros do have potential in promoting diversification of the world food supply and could make a significant

contribution if greater agronomic investment was made. The Root Crops program of the Food and Agriculture Organization of the United Nations (FAO) is attempting to address some of these issues, as is the work of the Australian Centre for International Agricultural Research (ACIAR) (Bradbury and Holloway 1988).

In the Pacific, people have a higher regard for taros than any of the other seven common starchy foods (Pollock 1992). Cassava outranks taros in terms of the tons per hectare produced, but that is because it is a good "safety" crop that will grow in poorer soils and can be harvested as needed when the more preferred starches are in short supply. Yet householders in most Pacific societies would not offer cassava to an honored guest or make it their contribution to a celebration; rather they would go to some lengths to procure *Colocasia taro* or *Dioscorea yams* or breadfruit for such purposes. In fact, *Colocasia taro*, yams, and breadfruit are at the very top of the list for everyday consumption in the Pacific and for exchanges and presentation at feasts. They are also the most expensive of the local foods on sale in the urban markets. The other three taros may be maintained as secondary or fallback foods, but the reputation of a rural family still rests in large part on its ability to produce a good supply of *Colocasia taro*, together with the other desirable crops, for self-maintenance (Pollock et al. 1989).

The taros (and other starch foods) form the major part of the daily diet of Pacific Island people living on their own land today, much as they have in the past. They are the main substance of daily intake, eaten once a day in the past, but now twice a day. Taros, and the other starches, provide the bulk of the food, the "real" food (*kakana dina* in Fijian), but are accompanied by a small portion of another foodstuff such as fish, coconut, or shellfish to form what we call a meal in English. If just one of them is eaten without the other, then people are likely to say that they have not eaten, because the two parts are essential to the mental, as well as physical, satisfaction that food confers (Pollock 1985).

Taro maintains this importance in the minds of contemporary Pacific Islanders living in metropolitan areas such as Wellington. The root may be expensive and hard to find, but these people make a great effort to obtain *Colocasia taro* for special occasions, such as a community feast, or for a sick Samoan or Tongan who may request a piece of taro to feel better.⁶

According to the accounts left by missionaries and other visitors to the Pacific in the nineteenth century, the amounts of taro (particularly *Colocasia taro*) consumed by Fijians and Tongans, for example, were prodigious. They especially noted the consumption patterns of chiefs, suggesting that all this taro was a cause of their obesity. We have less information, however, regarding the ordinary people's consumption (see Pollock 1992).

But in Tahiti, and probably elsewhere in the Pacific, food consumption generally varied from day to day and week to week. Europeans were amazed at how Tahitians could cross the very rugged interior of their island, going for four days with only coconut milk to drink, yet when food was available, they consumed very large amounts. Because food habits were irregular, one advantage of taro was that it made the stomach feel full for a long period of time.

Along with notions of routine introduced to the islands by missionaries and administrators came the concept of meals, which usually occur twice daily in rural areas. Taro might be eaten at both the morning and evening meals, and a schoolchild or an adult may carry a couple of slices of taro in a packed lunch. Indeed, schools in Niue are encouraging schoolchildren to bring their lunch in the form of local foods rather than bread and biscuits (Pollock 1983, field notes). Thus, today, an adult may consume a daily total of about 2 kg of *Colocasia taro* or other starch every day for 365 days of the year.

To a great extent such emphasis on local foodstuffs is the work of local Pacific food and nutrition committees, formed in the early 1980s, that have publicized the benefits of taro and other starches. But in urban areas of the Pacific, taros are scarce, and thus an expensive luxury food. In a Fijian or Samoan marketplace, four *Colocasia taros* (only enough to feed two adults for one meal) may sell for 5 or 6 dollars, with the other family members having to eat rice or cassava or *Xanthosoma taro* instead. Those promoting the use of local foods are endeavoring to bring down the price of taros. But to do so requires more agricultural input and other diversifications within the economy.

Colocasia taros are also an essential component of Pacific feasts where they take pride of place, alongside the pigs (used only at feasts), fish, turtle, or (today) beef. Early visitors to the Pacific were amazed at the immense walls of *Colocasia taros* and yams, topped off with pigs, that formed part of a presentation at a special occasion such as the investiture of a new chief in Fiji or Wallis in the 1860s. These food gifts were contributed by households closely associated with the community hosting the feast and were redistributed to those attending. A great amount of food had to be consumed at these feasts, as there was no means of preserving it (Pollock 1992).

Conversely, there were times when food was very scarce, as after a cyclone or a tidal wave, or during a drought. Such disasters witnessed the *Colocasia taro* plants damaged and rotted and the *Cyrtosperma* broken by the wind, so the people had to resort to dryland taro or *Alocasia taro* or other starches. In very severe cases (such as the devastating cyclone Val in December 1991 on Western Samoa), households had nothing but fallen coconuts and emergency foods, such as *Alocasia taro*, to rely on.

Exchanges of both planting material and of

harvested taros have constituted a method of adjusting such irregularity in food availability in the Pacific. Before the development of international aid in the 1960s and 1970s, taros and other starches were harvested in Tonga to aid neighbors and relatives in Wallis, Western Samoa, and Fiji. This process of exchange not only enabled families and villages to survive hard times, but it also cemented social relations between whole island nations. In addition, the process of exchange supported the development of a diversified gene pool of the various taros.

Cooking and Processing

All the taros must be cooked very thoroughly because of the oxalic acid crystals in the outer layer of the corm and in the leaves. Thorough cooking reduces the toxicity, and the earth oven allows whole taros to be covered and steamed on hot rocks for two hours or more. In most Pacific societies such an earth oven was made once a day, and in rural areas this is still the case. Boiling on a stove may be quicker, but it is more costly in fuel (Pollock 1992).

Pacific Island people today prefer taro cooked whole and then cut into slices for presentation to the household. Taro must be cooked as quickly as possible after harvesting to retain the best flavor and to avoid decay. Before cooking, each corm or stem of taro is carefully peeled, a process that can produce a skin irritation for those unaccustomed to it, again due to the oxalic acid crystals. The corms or stems are placed either in a coconut leaf basket or on banana leaves around the edge of the earth oven, with the fish (or pig if it is a feast) in the center, and the whole is covered first with leaves, then earth to allow the contents to steam. The oven is opened some two hours later. For special occasions, "puddings" may be made from grated taro mixed with coconut cream and baked in the earth oven.

One of the few societies to develop a processed form of taro was that in Hawaii, where fermented taro was eaten as *poi*. This was made by steaming, peeling, grinding, and straining the corms to yield a thick paste of 30 percent solids, known as "ready-to-mix" *poi*, or if more water was added to yield a thinner paste of 18 percent solids, known as "ready-to-eat" *poi*. Hawaiians refer to the thickness of *poi* as one-finger, two-finger, or three-finger *poi*. Either irrigated or dry-land Colocasia taro can be used for making *poi*, but different varieties of Colocasia taro are not mixed.

The thick paste ferments very rapidly due to lactobacilli fermentation, reaching an acidity level of 3.8 by the third day. Hawaiians would wrap the very thick paste, known as *'ai pa'i*, in *ti* leaves until needed. The addition of a little water to the desired portion was all that was required for serving highly esteemed *poi* to accompany fish or pork. The very thin paste, by contrast, lasts only three to four days unrefrigerated, and refrigerated *poi* becomes so rubbery that it is consid-

ered inedible (Moy and Nip 1983; Standal 1983; Pollock 1992).

Commercialization

Taros are sold whole and unprocessed in the Pacific. In Fiji, where the petioles are left attached to the corm, Colocasia taros are sold by the bundle of three or four tied together. In Tonga and Western Samoa, Colocasia taros are sold by the corm alone, but again in groups of four or more for a given price. The stems of Alocasia taros are sold by the piece, while the cormlets of Xanthosoma taro are sold by the basket, as are sweet potatoes and other root crops.

More of the crop is sold through middlemen in Fiji and Samoa, although producers themselves use family members as agents.⁷ Cyrtosperma taro is seldom sold in these larger markets, except in Tarawa, Kiribati, and Kolonia, Yap.

None of these root crops is very durable, so those marketing taro aim for quick sales. Damaged taros will deteriorate rapidly, hence great care is taken in both the harvesting process for market and in removing the tops in Tonga and Samoa to inflict as little damage to the corm as possible.

As early as 1880, Papeete in the Society Islands became a center for the redistribution of local produce (Pollock 1988). From such small waterside markets have grown the large market centers found around the tropical world today (some covering several acres). In each Pacific Island (and Caribbean Island) there is at least one such market in the urban center, and in larger islands, such as Fiji and Papua New Guinea, there are several markets in the various urban centers. These markets have grown in size and diversity over the last 20 years, as urban populations have increased. Only small amounts of taro are sold through supermarkets (Pollock 1988).

Out-migration of populations from the Pacific Islands (and the Caribbean) to metropolitan centers, such as Auckland, Wellington, Sydney, Honolulu, and Los Angeles, has also stimulated the overseas sale of taros, mainly Colocasia. The Tongan, Samoan, and Cook Islands populations are becoming sizable in those centers where demand for taro, mainly for celebratory occasions, has increased. Taro is available in urban markets, such as Otara in Auckland, and in vegetable shops, especially those where Polynesian communities are located. Prices are high, but families will make sacrifices to present some taro when needed to maintain family honor.

Before these outlets provided a steady supply, the various communities made private arrangements to import boxes of taro from their home islands. As a wider supply has become available and the communities have grown, each community has focused its demand on taro from its own island of origin, claiming that it tastes better. Samoans will track down stores that sell Samoan taro, whereas Tongans and

Rarotongans go in search of taros from their home islands. Island people themselves are acting more and more as the agents and middlemen, with the whole process promoting the production of taro varieties that will endure sea transport.

Taros are also imported in cooked form to New Zealand by returning residents. In Samoa or Niue, puddings are packed either in a chest freezer or a cardboard box and carried as part of the passenger's personal luggage. In New Zealand and Australia the families of the passenger then share in this produce "from home." Such is their social value that several hundred dollars may be spent in overweight luggage in order to transport local foods in this manner.

Another form of commercialization promoted by food and nutrition committees in various Pacific Islands is the use of taro (mainly *Colocasia*, both corm and leaves) along with other local foods, by hotels, to give tourists a new taste experience. Hawaii has long provided *luau* feasts for its visitors, which included a small portion of *poi* and pork, salmon, and coconut pudding. Now Fiji runs competitions in which chefs from leading hotels create recipes that make use of local foods, including taro. This practice, in turn, is leading to increased cooperation with the agriculture authorities to assist producers in regularizing production to supply the hotels.

In Hawaii, where processed taro has been marketed as *poi* for some 75 years, sales to Hawaiians and to the tourist hotels are supplemented by demand for *poi* in the mainland United States to help individuals suffering from allergies and digestive problems. As a consequence of this activity, Hawaii is the one place in the Pacific where taro plantations have become heavily commercialized and are run by companies rather than by family units.

Taro chips are now being manufactured in various centers around the Pacific. Local companies are selling their product, promoted by food and nutrition committees, in Fiji and Samoa with reasonable success. In Hawaii, entrepreneurial companies, such as Granny Goose Foods, are marketing taro chips alongside the traditional potato chips, thereby drawing taro into the lucrative snack industry.

In other parts of the tropical world, *Colocasia* taro may be processed into flour or flakes for commercial purposes. A product *Arvi* has been developed by the Central Food Technological Research Institute in Mysore, India, that consists of flour made from *Colocasia* taro. The corms are washed, peeled, and cut into slices, which are kept immersed in water overnight, then washed again and immersed for another three hours. The slices are blanched in boiling water for five minutes, then sun-dried before being ground into flour. A similar process has been used to make taro flour in Nigeria. The flour can be mixed with wheat flour for baking.

A process for making instant taro flakes has been tried in Taiwan and in Nigeria whereby smoke-dried

slices are stored away for later eating. Freezing taro has not been very successful, though a local variety was processed for freezing in Shanghai (Moy and Nip 1983). Taro leaves mixed with coconut cream, known in Samoa as *palusami*, have been canned with reasonable success, but the corm does not can well.

Nutritional Value

The nutritional value of taro has changed over the many years since it was first domesticated. Its users have selected plants that were less toxic, produced larger, less fibrous corms, and better suited their tastes. Such a selection process was facilitated by vegetative propagation, and many different cultivars were developed over time. However, a large proportion of these cultivars have been lost due to lack of interest in root crops by cereal-based colonial powers. Today the FAO and the South Pacific Commission are trying to preserve as many different cultivars in the Pacific as possible so as to increase the diversity of available food crops. *Colocasia* taro has many more different cultivars than the other three types of taro, indicating its preferred status and its longtime use as a food. The cultivars have different nutritional attributes.

The taro corms of the four different types vary slightly in their composition (see Table II.B.6.1 for details of composition of the four types of taro). All the corms consist mainly of starch and moisture and are high in fiber. They yield between 70 and 133 calories (or 255 and 560 kilojoules) per 100-gram portion, with *Alocasia* having the lowest range and *Xanthosoma* taro the highest. The amount of protein varies considerably from 1.12 percent to 2.7 percent depending on the type of taro, its geographical source, and the variety. The corms are also a good source of minerals, particularly calcium, for which *Cyrtosperma* taro is particularly notable (Standal 1982; Bradbury and Holloway 1988).

Taro leaves consist mainly of moisture and fiber. They are high in protein with a generally higher overall mineral content than the corms. It is only the young leaves of *Colocasia* taro that are eaten as a rule, although no difference in chemical composition has been found between leaves viewed as edible and those viewed as inedible (Bradbury and Holloway 1988). The use of the leaves as a wrapping in preparations, such as Samoan *palusami*, adds value to the diet on those special occasions when such a dish is served. Food and nutrition committees are trying to encourage the greater use of leaves, but they are not part of the traditional diet.

The fermented form of taro paste developed long ago by Hawaiians has been found to be a highly digestible product suitable for babies, adults with digestive problems, and those with allergies to cereals. The starch granules are small enough to pass readily into the digestive system. This attribute has led to the commercialization of *poi* (Standal 1983).

Table II.B.6.1. *Nutritional value of the four types of taros*

	Energy		Protein (g)	Fat (g)	C.H.O. (g)	Ca (mg)	Iron (mg)	Vit. A (µg)	Thiamine (mg)	Riboflavin (mg)	Niacin (mg)	Vit. C (mg)	Waste
	(Kcal)	(MJ)											A.C. (%)
Taro (Colocasia)	113	0.47	2.0	-	26.0	25	1.0	-	0.100	0.03	1.0	5	20
Taro, giant (Alocasia)	70	0.29	0.6	0.1	16.9	152	0.5	-	0.104	0.02	0.4	-	-
Taro, swamp (Cyrptosperma)	122	0.51	0.8	0.2	29.2	577	1.3	-	0.027	0.11	1.2	-	-
Xanthosoma	133	-	2.0	0.3	31.0	20	1.0	-	1.1	0.03	0.5	10.1	?

Clearly, taro has considerable merits as a food. It is readily cooked in an earth oven with minimal equipment, or it can be boiled or baked on a stove. It provides a high-bulk foodstuff rich in fiber, with acceptable amounts of vegetable protein and calcium. There is enough variety of cultivars to yield different tasting corms (if taste is an important consideration). But these merits have not been recognized widely enough, an issue the FAO Root Crops Program in the South Pacific is attempting to rectify through agricultural development (Sivan 1984; Jackson and Breen 1985). Simultaneously, food and nutrition committees, through their promotion of local foods, are endeavoring to counter the colonial legacy that bread is best.

Summary

Taro has evolved as a food over several thousand years, as people in tropical areas have selected attributes that suit their needs. Those needs included both consumption and production factors, as well as processing techniques. In the Pacific area, where the taros are most widely used, the people have relied heavily on three forms, Colocasia, Alocasia, and Cyrptosperma, along with the other starches such as yams, breadfruit, and bananas as the main elements in their daily diets, eaten together with a small accompanying dish. Xanthosoma taro has been added to this inventory in the last 200 years, as it will grow in poor soils and can be less acid.

Vegetation propagation allowed a high degree of selectivity. Factors including the taste of the corm and its size, color, moisture, and acidity have determined over time which setts were replanted and which were discarded.

Most taro has been grown in dryland conditions. The selection of varieties of Colocasia taro that would grow in water is a further development, as is the very specialized technique for raising Cyr-

ptosperma taro on atolls where the salinity of the water is a problem.

Little development has taken place to diversify the edible product. The corms are peeled and cooked in an earth oven by steaming for a couple of hours and are then served in slices. More recently, boiling has been introduced, but it gives a less acceptable flavor.

Ongoing development of the taros was curtailed, to some extent, by colonial Europeans whose preferred food was bread. Taros and other root crops were considered by these newcomers as a mark of the backward nature of these societies, and the colonists introduced crops of a commercial nature, such as cotton, vanilla, sugar cane, and, more recently, coffee and cocoa. These crops were planted on the best land, and taros were relegated to less desirable areas. The result has been not only a loss of many varieties of taro formerly used but also a scarcity of taros for sale in the markets today over and above those needed for household supply.

Only during the last decade of the twentieth century have the root crops, including taro, merited the attention of agricultural specialists. The worldwide pressure for a more differentiated crop base than just the seven basic food crops has led to programs such as the FAO Root Crops Program and ACIAR's identification of the potential of root crops in the South Pacific. With political independence in the 1960s and 1970s, small nations in the tropics have seen the need to become more self-reliant by reducing their high food import bills. The former importance of the taros has been recognized, and these countries are now taking steps to reestablish them agronomically and economically as a key local crop. The recognition of the importance to health of dietary fiber adds another dimension to taro's desirability. Exports of taro to migrants in metropolitan areas have stimulated the need for particular farming expertise as

well as the development of marketing and processing techniques.

Taro has survived a major hiatus in the nineteenth and twentieth centuries that might have seen it eliminated as a crop or dismissed as one of backward, underdeveloped tropical countries. But cereals, even rice, will not grow readily in many of these tropical areas, whereas the taros are a flexible crop that suits shifting cultivation so that farmers can vary the size of their crops from month to month depending on demand. Nutritionally, taro is very good, especially when complemented with fat from fish or pork. Given agronomic support, taro has great potential for further contributions to the world food supply, and finally, it is a crop that has endured thanks to people's strong preference for it as their traditional food.

Nancy J. Pollock

Notes

1. It is ironic that rice has been introduced into modern-day Pacific diets as an emergency foodstuff that is easily transferred from metropolitan countries as a form of food aid to assist cyclone-stricken nations, such as Samoa in 1991. As such, it forms a substitute for locally grown taro, which is badly affected by salt inundation and wind breaking off the leaves, thus causing the corms to rot.
2. See Petterson (1977: 177) for a map of the spread of Xanthosoma taro around central and northern South Africa in contrast with the spread of Colocasia taro in the same area.
3. See Pollock (1992) for a listing of the importance of various starch staples in Pacific societies.
4. See Barrau (1965: 69) for a map showing its origins and distribution in Southeast Asia and the Pacific.
5. Petterson (1977: 178); see also the map on p. 177.
6. See Pollock et al. 1989 for preferences and consumption patterns of taros and other foods by those Samoans living away from their home islands in Wellington, New Zealand.
7. See Chandra (1979) for a detailed discussion of marketing root crops.

Bibliography

- Barrau, J. 1961. *Subsistence agriculture in Polynesia and Micronesia*. B.P. Bishop Museum Bulletin No. 223.
1965. L'humide et le sec. *Journal of the Polynesian Society* 74: 329-46.
1975. The Oceanians and their food plants. In *Man and his foods*, ed. C. Earle Smith, Jr., 87-117. Tuscaloosa, Ala.
- Bellwood, P. 1980. Plants, climate and people. In *Indonesia, Australia perspectives*, ed. J. J. Fox, 57-74. Canberra, Australia.
- Bradbury, H., and W. D. Holloway. 1988. *Chemistry of tropical root crops*. Canberra, Australia.
- Chandra, S. 1979. Root crops in Fiji, Part I. *Fiji Agricultural Journal* 41: 73-85.

- Chang, K. C. 1977. *Food in Chinese culture*. New Haven, Conn.
- de la Pena, R. 1983. Agronomy. In *Taro*, ed. J.-K. Wang, 169-70. Honolulu.
- Dignan, C. A., B. A. Burlingame, J. M. Arthur, et al., eds. 1994. *The interim Pacific Islands food composition tables*.
- Gorman, C. 1977. A priori models and Thai history. In *The origins of agriculture*, ed. C. A. Reed, 321-56. Mouton, N.S., Canada.
- Handy, E. S. C., and Willowdean Handy. 1972. *Native planters in old Hawaii*. Honolulu.
- Holo, T. F., and S. Taumoeofolau. 1982. The cultivation of *alocasia macrorrhiza* (L.) schott. In *Taro cultivation in the Pacific*, ed. M. Lambert, 84-7. Noumea, New Caledonia.
- Hutterer, Karl. 1983. The natural and cultural history of South East Asian agriculture. *Anthropos* 78: 169-201.
- Jackson G. V. H., and J. A. Breen. 1985. *Collecting, describing and evaluating root crops*. Noumea, New Caledonia.
- Kirch, Patrick V. 1985. *Feathered gods and fishbooks*. Honolulu.
- Lambert, Michel. 1982. *Taro cultivation in the South Pacific*. Noumea, New Caledonia.
- Luomala, Katharine. 1974. The Cyrtosperma systemic pattern. *Journal of the Polynesian Society* 83: 14-34.
- Massal, E., and J. Barrau. 1956. *Food plants of the South Sea Islands*. Noumea, New Caledonia.
- Mitchell, W. C., and Peter Maddison. 1983. Pests of taro. In *Taro*, ed. J.-K. Wang, 180-235. Honolulu.
- Moy, J. H., and W. Nip. 1983. Processed foods. In *Taro*, ed. J.-K. Wang, 261-8. Honolulu.
- Murdock, G. P. 1960. Staple subsistence crops of Africa. *Geographical Review* 50: 523-40.
- Norman, M. J. T., C. J. Pearson, and P. G. E. Searle. 1984. *The ecology of tropical food crops*. London.
- Ooka, J. J. 1983. Taro diseases. In *Taro*, ed. J.-K. Wang, 236-58. Honolulu.
- Pawley, A., and R. Green. 1974. The proto-oceanic language community. *Journal of Pacific History* 19: 123-46.
- Petterson, J. 1977. Dissemination and use of the edible aroids with particular reference to Colocasia (Asian Taro) and Xanthosoma (American Taro). Ph.D. thesis, University of Florida.
- Plucknett, D. L. 1976. Edible aroids. In *Evolution of crop plants*, ed. N. W. Simmonds, 10-12. London.
1983. Taxonomy of the genus *Colocasia*. In *Taro*, ed. J.-K. Wang, 14-19. Honolulu.
- Pollock, Nancy J. 1983. Rice in Guam. *Journal of Polynesian Society* 92: 509-20.
1985. Food concepts in Fiji. *Ecology of Food and Nutrition* 17: 195-203.
1988. The market place as meeting place in Tahiti. In *French Polynesia*, ed. Nancy J. Pollock and R. Cromcombe, Suva, Fiji.
1990. Starchy food plants in the Pacific. In *Nga Mabi Maori O te Wao Nui a Tane, Contributions to an international workshop on ethnobotany*, ed. W. Harris and P. Kapoor, 72-81. Canterbury.
1992. *These roots remain*. Honolulu.
- Pollock, N. J., A. Ahmu, S. Asomua, and A. Carter. 1989. Food and identity: Food preferences and diet of Samoans in Wellington, New Zealand. In *Migrations et identité, actes du colloque C.O.R.A.I.L.*, Publications de l'Université Française du Pacifique, Vol. 1. Noumea, New Caledonia.
- Purseglove, J. W. 1972. Tropical crops. In *Monocotyledons I*, 58-75. New York.
- Seeman, B. 1862. *Viti*. London.

- Sivan, P. 1984. Review of taro research and production in Fiji. *Fiji Agricultural Journal* 43: 59-68.
- Spriggs, M. 1982. Taro cropping systems in the South East Asian Pacific region. *Archeology in Oceania* 17: 7-15.
- Standal, B. 1982. Nutritional value of edible aroids (Araceae) grown in the South Pacific. In *Taro cultivation in the South Pacific*, ed. M. Lambert, 123-31. Noumea, New Caledonia.
1983. Nutritive value. In *Taro*, ed. J.-K. Wang, 141-7. Honolulu.
- Tang, C., and W. W. Sakai. 1983. Acridity of taro and related plants in Araceae. In *Taro*, ed. J.-K. Wang, 148-64. Honolulu.
- Vickers, M., and V. Untaman. 1982. The cultivation of taro *cyrtosperma chamissonis* schott. In *Taro cultivation in the South Pacific*, ed. M. Lambert, 90-100. Noumea, New Caledonia.
- Weightman, B., and I. Moros. 1982. The cultivation of taro *xanthosoma* sp. In *Taro cultivation in the South Pacific*, ed. M. Lambert, 74-83. Noumea, New Caledonia.
- Yen, D. E. 1980. The South East Asian foundations of Oceanic agriculture. *Journal de la Societe des Oceanistes* 66-7: 140-7.



II.C

Important Vegetable Supplements

II.C.1 Algae

Algae are eukaryotic photosynthetic micro- and macroorganisms found in marine and fresh waters and in soils. Some are colorless and even phagotrophic or saprophytic. They may be picoplankton, almost too small to be seen in the light microscope, or they could be up to 180 feet long, such as the kelp in the kelp forests in the Pacific Ocean.

Algae are simple, nucleated plants divided into seven taxa: (1) Chlorophyta (green algae), (2) Charophyta (stoneworts), (3) Euglenophyta (euglenas), (4) Chrysophyta (golden-brown, yellow-green algae and diatoms), (5) Phaeophyta (brown algae), (6) Pyrrophyta (dinoflagellates), and (7) Rhodophyta (red algae). A taxon of simple, nonnucleated plants (prokaryotes) called Cyanobacteria (blue-green bacteria) is also included in the following discussion as they have a long history as human food.

Algae are eaten by many freshwater and marine animals as well as by several terrestrial domesticated animals such as sheep, cattle, and two species of primates: *Macaca fuscata* in Japan (Izawa and Nishida 1963) and *Homo sapiens*. The human consumption of algae, or phycophagy, developed thousands of years ago, predominantly among coastal peoples and, less commonly, among some inland peoples. In terms of quantity and variety of species of algae eaten, phycophagy is, and has been, most prevalent among the coastal peoples of Southeast Asia, such as the ancient and modern Chinese, Japanese, Koreans, Filipinos, and Hawaiians.

History and Geography

The earliest archaeological evidence for the consumption of algae found thus far was discovered in ancient middens along the coast of Peru. Kelp was found in middens at Pampa, dated to circa 2500 B.C. (Moseley 1975); at Playa Hermosa (2500–2275 B.C.); at Concha (2275–1900 B.C.); at Gaviota (1900–1750 B.C.); and at

Ancon (1400–1300 B.C.) (Patterson and Moseley 1968). T. C. Patterson and M. E. Moseley (1968) believe that these finds indicate that marine algae were employed by the ancient Peruvians to supplement their diets.

Other types of seaweeds were also found in middens at Aspero, Peru, and dated to 2275 to 1850 B.C. by Moseley and G. R. Willey (1973) and at Asia, Peru, and dated to 1314 B.C. (Parsons 1970). Additionally, unidentified algae were found in middens at numerous sites, among them Padre Aban and Alto Salaverry (2500–1800 B.C.); Gramalote, Caballo Muerte (2000–1800 B.C.); Cerro Arena, Moche Huacas (200 B.C. to A.D. 600); and Chan Chan (A.D. 1000–1532) (Pozorski 1979; Raymond 1981).

Furthermore, much evidence exists to indicate a marine algae presence in ancient Peru. The base of the temples at Las Haldas, for example, which date circa 1650 B.C., contained quantities of seaweed and shellfish (Matsuzawa 1978). Small stalks of a seaweed were found in a Paracas mummy bundle (Yacovleff and Muelle 1934: 134), and the giant kelp *Macrocystis humboldtii* was pictured on an ancient Nazca vase (Yacovleff and Herrera 1934–5). In recent times, the native peoples of the Andean highlands have retained the right to come down to the coast, gather and dry seaweed, and transfer it to the mountains, where algae has great value and can be used in place of money. Used as a condiment to flavor soups and stews, dried seaweed minimizes the ravages of hypothyroidism, which is endemic in the Andes (Aaronson 1986). Early visitors reported that dried seaweed was also eaten with vinegar after dinner and sold in the marketplace as a kneaded dry product (Cobo 1956). The cyanobacteria *Nostoc* spp. (called *cushuro*, *llucllucha*, or *kochayuyo*) grow in Andean lakes and ponds and are also presently used as food (Aldave-Pajares 1965–66; Gade 1975; Browman 1981; and Table II.C.1.1), as they were in early Spanish colonial times (Cobo 1956) and, possibly, in Inca times as well (Guaman Poma de Ayala 1965–6).

Table II.C.1.1. *Algae and blue-green bacteria eaten in contemporary Chile and Peru*

Species	Common name	Eaten as ^a
Cyanobacteria		
<i>Nostoc commune</i>	<i>cusburo</i> ^b	<i>picantes</i>
	<i>elulluchba</i> ^b	-
	<i>kocha-yuya</i> ^b	-
<i>Nostoc parmeloides</i>	<i>llullcha</i> ^b	-
<i>Nostoc pruniforme</i>	<i>cusburo</i> , <i>cusuro</i> ^b	<i>chupe</i>
	<i>crepito</i> ^b	<i>loco</i>
	<i>unrupa macho</i> ^b	<i>picantes</i>
	<i>macha-masba</i> ^b	<i>mazamorro</i>
<i>Nostoc sphaericum</i>	<i>casburo</i> ^b	<i>chupe</i>
	<i>cussuro</i> , <i>cusuro</i> ^b	<i>loco</i>
	<i>ururupsba</i> ^b	<i>mazamorro</i>
	<i>rachapa</i> ^b	<i>picantes</i>
<i>Nostoc verrucosum</i>	<i>murmunta</i> ^b	-
Chlorophyta		
<i>Codium fragile</i>	<i>tercioelo</i> ^c	-
<i>Monostroma</i> sp.	<i>yuyo de río</i> ^b	noodle soup
	<i>lebuga de mar</i> ^b	-
	<i>lechuquita de río</i> ^b	-
<i>Ulva fasciata costata</i>	<i>lechuquita de río</i> ^b	<i>picante</i>
<i>Ulva lactuca</i>	<i>luce verde</i> ^d	-
<i>Ulva pappenfussii</i>	<i>cobayuyo</i> ^b	-
Phaeophyta		
<i>Durvillea antarctica</i>	<i>cobayuyo</i> ^{c,d}	-
<i>Lessonia nigrescens</i>	<i>tinilbue</i> ^c	-
<i>Macrocystis integrifolia</i>	<i>buiro</i> ^d	-
<i>Macrocystis pyrifera</i>	<i>buiro</i> ^d	<i>ceviche</i>
Rhodophyta		
<i>Gigartina chamissoi</i>	<i>cobayuyo</i> ^b	-
	<i>chicorea del mar</i> ^c	-
<i>Gigartina glomerata</i>	<i>cobayuyo</i> ^b	<i>ceviche</i>
	<i>yuyo</i> ^b	<i>picantes</i>
	<i>mococho</i> ^b	soup
<i>Gigartina paitensis</i>	<i>mococho</i> ^b	soup
	<i>cobayuyo</i> ^b	<i>picantes</i>
<i>Gracillaria verrucosa</i>	<i>petillo</i> ^c	-
<i>Iridaea boryana</i>	<i>luga</i> ^c	-
<i>Porphyra columbina</i>	<i>luce</i> ^c	-
<i>Porphyra leucosticta</i>	<i>cobayuyo</i> ^b	<i>picantes</i>
<i>Prionitis decipiens</i>	<i>piscuchaqui</i> ^b	<i>picantes</i>
<i>Rhodoglossum denticulatum</i>	<i>cobayuyo</i> ^b	<i>ceviche</i>
		<i>picantes</i>
		soup

^aCommon names for Peruvian food from Polo (1977).^bPolo (1977).^cDillehay (1989).^dMasuda (1985).

Note: *Ceviche* is a soup with algae, small pieces of fish, lemon, and hot pepper.

Chupe is a soup with milk, algae, eggs, potatoes, and cheese.

Cobayuyo is from the Quecha *cocba* (lagoon or pond) and *yuyo* (herb or vegetable).

Cusburo is from the Quecha for wavy.

Pachayuyo is from the Quechua *pacho* (soil) and *yuyo* (vegetable).

Picantes is a stew made of algae, pieces of fish, potatoes, and hot pepper.

Locro is a maize and meat stew with algae.

Mazamorro is a stew made with pieces of algae and other ingredients.

Moving north in the Americas, *Spirulina maxima*, or *Spirulina geitlerii* (a blue-green bacterium known as *tecuitlatl* [stone excrement] in Nahuatl, the Aztec language), has been eaten in the Valley of Mexico since the beginning of the Spanish colonial period (c. 1524) and was consumed even prior to Aztec times (Furst 1978). Other cyanobacteria, such as *Phormidium tenue* and *Chroococcus turgidus* (called *cocolin* and *Nostoc commune* [*amoxtle* in Nahuatl]), are gathered for consumption from the lakes and ponds of the Valley of Mexico and, very likely, have been since time immemorial (Ortega 1972).

In Africa, another species of cyanobacterium, *Spirulina platensis*, grows abundantly in Lake Chad and is collected, dried, and made into a sauce. It is widely consumed by the Kanembu people of Chad (Leonard and Compère 1967; Delpeuch, Joseph, and Cavalier 1975).

In China, the earliest reference to algae as food occurs in the *Book of Poetry* (800–600 B.C.) (Chase 1941), and Wu's *Materia Medica* indicates that the seaweed *Ecklonia* was utilized as food and medicine as early as 260 to 220 B.C. Another type of seaweed, *Porphyra*, was also used as food according to the *Qiminyaosu* (A.D. 533–44).

Gloiopeltis furcata has been collected in southern Fujian Province since the Sung Dynasty (A.D. 960–1279) (Tseng 1933), and C.-K. Tseng (1987) states that *Laminaria japonica* (*baidi*) has been eaten there for about 1,000 years. Several other types of algae were also consumed, according to the *Compendium of Materia Medica* (*Bencaigangmu*) of the Ming Dynasty (1368–1644), edited by L. Shizhan (1518–93). These included: *Cladophora* spp., *Codium fragile*, *Ecklonia kurone*, *Enteromorpha prolifera*, *Enteromorpha* spp., *Euchema muricatum*, *Gelidium divaricatum*, *Gloiopeltis furcata*, *Gracillaria verrucosa*, *Laminaria japonica*, *Monostroma nitidum*, *Nostoc sphaeroides*, *Prasiola sinensis*, *Porphyra* sp., and *Ulva lactuca*.

The cyanobacterium *Nostoc commune* has been eaten in China for the last 400 years (Chu and Tseng 1988), and a related species, *Nostoc coeruleum*, was served in this century at a dinner given by a mandarin for a French ambassador (Tilden 1929; Montagne 1946–7). According to Tseng (1990), the large-scale cultivation of the seaweed *Gloiopeltis furcata* began in Fujian Province during the Sung Dynasty.

In Japan, the eating of algae is also an ancient practice. Seaweed was apparently eaten by the early inhabitants of Japan, as it has been found with shells and fish bones at human sites in the Jomon period (10,500–300 B.C.) and the Yayoi period (200 B.C. to A.D. 200) (Nisizawa et al. 1987). In A.D. 701, the emperor established the Law of Taiho in which seaweeds (*Gelidium*, *Laminaria*, *Porphyra*, and *Undaria* spp.) were among the marine products paid to the court as a tax (Miyashita 1974).

The blue-green bacterium *Nostoc verrucosum*, currently known as *ashitsuki nori*, was mentioned

in *Man Yo Shu* by Yakomichi Otomi in the oldest anthology of 31-syllable odes dating to A.D. 748. Ode number 402 translates: "Girls are standing on the shores of the Wogami-gawa River in order to pick up *asbitsuki nori*." According to the *Wamyosho* (the oldest Chinese-Japanese dictionary in Japan), 21 species of marine algae (brown, green, and red) were eaten by Japanese during the Heiam era (A.D. 794–1185) (Miyashita 1974). In *The Tale of the Genji*, written by Murasake Shikibu (A.D. 978), a marine alga, *Codium fragile*, is mentioned as a food (Hiroe 1969). The Greeks and Romans apparently disliked algae and, seemingly, made no use of them as human food, although they were used as emergency food for livestock. Virgil (70–19 B.C.) called algae *vilior alga* (vile algae), and Horace (65–8 B.C.) seems to have shared his opinion (Newton 1951; Chapman 1970).

The seaweed *Rhodomenia palmata* was eaten in Iceland as early as A.D. 960, according to the *Egil Saga* (Savageau 1920). *Alaria esculenta* (bladderlocks) was (and still is) consumed in Scotland, Ireland, Iceland, Norway, and the Orkney Islands, where it is called alternatively "honey-ware," "mirkles," and "murlins." *Laminaria digitata* ("tangle") is also eaten in Scotland, and *Rhodomenia palmata* ("dulse") is used as food in Iceland, Scotland, and around the Mediterranean, where it is an ingredient in soups and

ragouts. Similarly, *Laurencia pinnatifida* ("pepper dulse") is eaten in Scotland, and *Porphyra laciniata* ("purple laver") is used as a condiment in the Hebrides (Johnston 1970).

Algae and Cyanobacteria as Human Food Today

Algae (Chlorophyta, Phaeophyta, and Rhodophyta) and cyanobacteria are now consumed in all countries that possess marine coasts as well as in countries where algae are abundant in lakes, streams, ponds, and rivers. Consumers range from surviving Stone Age peoples to modern hunter-gatherers, to agricultural folk, to industrial peoples.

Algae are used as foods in a wide variety of ways. They are served raw in salads and pickled or fermented into relish. They make a fine addition to soups, stews, and sauces, and are used as condiments. Algae are also roasted, employed as a tea, and served as a dessert, a sweetmeat, a jelly, or a cooked vegetable (see Table II.C.1.2 for countries in which algae are eaten as food and the form of the food).

In industrialized countries, algal products like agar, alginates, and carrageenans are extracted from some seaweeds and used to replace older foods or to create new foods or food combinations.

Table II.C.1.2. *Algae eaten by humans now and in the past*

Species name	Country	Local name	Use	Reference
Cyanophyta				
<i>Apbanotbece sacrum</i>	Japan	-	Food	Zaneveld 1950; Watanabe 1970
<i>Brachytrichia quoyi</i>	China	-	Food	Chu and Tseng 1988
	Japan	-	Food	Watanabe 1970
<i>Chroococcus turgidus</i>	Mexico	<i>Cocol de agua</i>	Food	Ortega 1972
<i>Nostoc coeruleum</i>	China	-	Food	Montagne 1946–7
<i>Nostoc commune</i>	Peru	<i>cushuro</i>	Food	Aldave-Pajares 1969
	Japan	<i>kamagawa-nori</i>	Food	Watanabe 1970
	Java	<i>djamurbatu</i>	Food	Zaneveld 1950
	Mongolia	-	Food	Elenkin 1931
	Mexico	<i>amoxtle</i>	Food	Ortega 1972
	Bolivia	-	Food	Lagerheim 1892
	Ecuador	-	Food	Lagerheim 1892
<i>Nostoc commune</i> var. <i>flagelliforme</i>	China	<i>fa-ts'ai</i>	Food	Jassley 1988
<i>Nostoc edule</i> = <i>pruniforme</i>	China, Mongolia, Soviet Union	-	Soup	Johnston 1970
	Peru	<i>cushuro</i>	Food	Aldave-Pajares 1969
<i>Nostoc ellipsosporum</i>	Ecuador	-	Food	Lagerheim 1892
<i>Nostoc parmeloides</i>	Peru	-	Food	Aldave-Pajares 1985
<i>Nostoc spbaericum</i>	Peru	<i>cushuro</i>	Food	Aldave-Pajares 1969
<i>Nostoc verrucosum</i>	Peru	<i>cushuro</i>	Food	Aldave-Pajares 1969
	Thailand	-	Food	Smith 1933
	Japan	-	Food	Watanabe 1970
<i>Nostoc</i> sp.	Fiji	-	Food	Wood 1965
<i>Nostocopsis lobatus</i>	China	-	Food	Chu and Tseng 1988
<i>Nostocopsis</i> sp.	Thailand	-	Soup, dessert	Lewmanomont 1978
<i>Oscillatoria</i> spp.	Java	<i>keklap</i>	Food	Zaneveld 1950
<i>Phormidium tenue</i>	Mexico	<i>cocol de agua</i>	Food	Ortega 1972
<i>Phyloderma sacrum</i>	Japan	<i>suizenji-nori</i> , <i>kotobuki-nori</i>	Food	Watanabe 1970
<i>Spirulina maxima</i>	Mexico	<i>tecuitlatl</i>	Food	Clément, Giddey, and Merzi 1967
<i>Spirulina platensis</i>	Chad	<i>die</i>	Sauce	Dangeard 1940

(continued)

Table II.C.1.2. (Continued)

Species name	Country	Local name	Use	Reference
Charophyta				
<i>Chara</i> spp.	Peru	-	Food	Browman 1980
Chlorophyta				
<i>Caulerpa</i> sp.	New Caledonia	-	Baked, raw	Barrau 1962
	New Hebrides	-	Salad	Massal and Barrau 1956
	Polynesia	<i>limu, limu</i>	Salad	Massal and Barrau 1956
<i>Caulerpa peltata</i>	Malaysia	<i>lata</i>	Salad	Zaneveld 1950
	Java	<i>lata</i>	Salad	Zaneveld 1955
	Philippines	-	Salad	Zaneveld 1955
	Indonesia	-	Salad,	
			sweetmeat	Subba Rao 1965
<i>Caulerpa racemosa</i>	Malaysia	<i>letato</i>	Dessert	Zaneveld 1955
	Philippines	<i>ararucip lai-lai</i>	Salad	Galutira and Velasquez 1963
	Indonesia	-	Raw,	
			sweetmeat	Subba Rao 1965
	Guam	<i>limu fuafua</i>	Relish	Subba Rao 1965
	Melanesia	-	-	Massal and Barrau 1956
<i>Caulerpa serrulata</i>	Malaysia	-	Food	Subba Rao 1965
	Celebes	-	Food	Zaneveld 1955
	Singapore	-	Salad	Zaneveld 1955
	Philippines	<i>gal galacgac</i>	Food	Zaneveld 1950
<i>Chaetomorpha</i> sp.	Malaysia	<i>lumut laut</i>	Food	Zaneveld 1950
<i>Chaetomorpha antennina</i>	Hawaii	<i>limu butuito, limu ilic, limumami</i>	Food	Zaneveld 1959
	Malaysia	<i>lumut-laut</i>	Raw	Zaneveld 1950
<i>Chaetomorpha crassa</i>	Philippines	<i>kauat-kauat</i>	Sweetmeat	Zaneveld 1950
<i>Chaetomorpha javanica</i>	Indonesia	<i>lumut laut</i>	Food	Zaneveld 1950
<i>Chaetomorpha tomentosum</i>	Malaysia	<i>susu-lopek, laur-laur</i>	Raw	Zaneveld 1950
<i>Cladophora</i> sp.	Peru	-	Food	Browman 1980
<i>Codium</i> sp.	Philippines	<i>pocpoclo</i>	Salad	Galutira and Velasquez 1963
	Samoa	-	Raw, baked	Barrau 1962
<i>Codium fragile</i>	Japan	<i>miru</i>	Soup, sauce	Chapman and Chapman 1980
	China	<i>shulsong</i>	Food	Tseng 1983
<i>Codium intricatum</i>	Philippines	<i>pocpoclo</i>	Salad, cooked	
			vegetable	Velasquez 1972
<i>Codium muelleri</i>	Hawaii	<i>limu aalaula</i>	Salad	Zaneveld 1955
	Philippines	<i>pocpoclo</i>	Salad	Zaneveld 1955
<i>Codium papillatum</i>	Philippines	<i>pocpoclo</i>	Salad, cooked	
			vegetable	Velasquez 1972
<i>Codium tomentosum</i>	Malaysia	<i>susu-lopek, laur-laur</i>	Raw	Zaneveld 1950;
				Subba Rao 1965
<i>Enteromorpha</i> sp.	China	<i>bu-t'ai</i>	Condiment	Tseng 1933
	Hawaii	<i>limu eleele</i>	Salad	Abbott 1978
	New Caledonia	-	Raw, baked	Barrau 1962
	Philippines	<i>lumot</i>	Salad	Galutira and Velasquez 1963
	Northwest North America	-	Food	Madlener 1977
<i>Enteromorpha clatrata</i>	China	<i>taitiao</i>	Food	Tseng 1983
<i>Enteromorpha compressa</i>	Malaysia	-	Food	Zaneveld 1955
	Philippines	<i>lumot</i>	Salad	Velasquez 1972
<i>Enteromorpha flexuosa</i>	Hawaii	<i>limu eleele</i>	Salad	Chapman and Chapman 1980
<i>Enteromorpha intestinalis</i>	Canada	-	Food	Turner 1974
	Malaysia	-	Food	Zaneveld 1950
	Philippines	<i>lumot</i>	Salad	Zaneveld 1955
<i>Enteromorpha plumosa</i>	Philippines	<i>lumot</i>	Salad	Galutira and Velasquez 1963
<i>Enteromorpha prolifera</i>	Hawaii, Malaysia, China	<i>limu eleele</i>	Salad	Zaneveld 1950
<i>Enteromorpha tubulosa</i>	China	<i>bu-t'ai</i>	Food	Tseng 1983
<i>Monostroma</i> sp.	Peru	-	Soup	Polo 1977
	China	<i>chiao-mo</i>	Condiment	Tseng 1933
<i>Monostroma nitidum</i>	China	-	Condiment	Xia and Abbott 1987
<i>Monostroma guaternaria</i>	Peru	-	Food	Aldave-Pajares 1985
<i>Oedogonium</i> sp.	India	-	Food	Tiffany 1958
<i>Prasiola japonica</i>	Japan	<i>kawa-nori, daiyagawa-nori, nikko-nori</i>	Food	Namikawa 1906;
				Skvortzov 1919-22;
				Watanabe 1970
<i>Prasiola yunnanica</i>	China	-	Food	Jao 1947; Jassley 1988
<i>Spirogyra</i> sp.	Burma	-	Food	Biswas 1953
	Canada	-	Food	Turner 1974
	Indochina	-	Food	Léonard and Compère 1967
	Thailand	-	Soup, salad	Lewmanomont 1978

Table II.C.1.2. (Continued)

Species name	Country	Local name	Use	Reference
Chlorophyta				
<i>Ulothrix flacca</i>	China	-	Vegetable	Xia and Abbott 1987
<i>Ulva</i> sp.	China	-	Food	Chapman and Chapman 1980
	Japan	<i>awosa, aosa</i>	Garnish	Chapman and Chapman 1980
<i>Ulva conlobata</i>	China	-	Tea	Xia and Abbott 1987
<i>Ulva fasciata</i>	Hawaii	<i>limu pabapaba</i>	Food	Schönfeld-Leber 1979
	Peru	<i>cocbayuyo</i>	Food	Masuda 1981
	China	-	Tea	Xia and Abbott 1987
<i>Ulva lactuca</i>	Canada (Bella Coola, Haida, Lillooet Indians)	-	Food	Turner 1974
	Chile	<i>luche</i>	Food	Ohni 1968
	China	<i>Hai ts'ai</i>	Soup, salad, vegetable	Tseng 1933
	New Zealand (Maori)		Soup, stew, vegetable	Brooker, Combie, and Cooper 1989
	Peru	<i>cocbayuyo</i>	Food	Masuda 1981
	Philippines	<i>gamgamet</i>	Salad, cooked vegetable	Velasquez 1972
<i>Ulva lactuca</i>	United States, California (Pomo and Kashoia Indians)	<i>sibtono</i>	Flavoring	Goodrich, Lawson, and Lawson 1980
	Hawaii	<i>limu pakcaea</i>	Soup, salad, garnish	Zaneveld 1955
	Washington (Makah)	<i>kalkatsup</i>	Food	Gunther 1945
Phaeophyta				
<i>Alaria crassifolia</i>	Japan	<i>chigaiso</i>	Food	Subba Rao 1965
<i>Alaria esculenta</i>	Alaska (Indian)	-	Food	Porsild 1953
	Siberia (Nivkhi)	-	Food	Eidlitz 1969
	Iceland, Ireland, Orkney Islands, Norway, Scotland	-	Food	Johnston 1970
<i>Alaria pylati</i>	Greenland (Inuit)	<i>kjipilasat</i>	Food	Hoygaard 1937
	Greenland (Angmagsalik)	<i>suvdluitsit</i>	Food	Ostermann 1938
	Siberia	<i>me'cgomei</i>	Food	Eidlitz 1969
<i>Artbrothammus bifidus</i>	Japan	<i>mekoasbi-</i> <i>kombu chigaiso</i>	Food	Subba Rao 1965
<i>Artbrothammus kurilensis</i>	Japan	<i>chisbimanekoasbi</i>	Food	Subba Rao 1965
<i>Ascophyllum nodosum</i>	Greenland	<i>miserarnat</i>	Food	Hoygaard 1937
<i>Chnoospora pacifica</i>	Indochina	<i>rau ngoai</i>	Salad, relish	Zaneveld 1955
<i>Chorda filum</i>	China	-	Food	Tseng 1983
	Japan	-	Salad	Chapman and Chapman 1980
<i>Dictyopteris plagiogramma</i>	Hawaii	<i>limu lipoa</i>	Food	Chapman and Chapman 1980
<i>Dictyopteris repens</i>	Easter Island	<i>auke</i>	Food	Ohni 1968
<i>Dictyota</i> sp.	Indonesia	-	Food	Michanek 1975
<i>Dictyota acutiloba</i>	Hawaii	-	Food	Chapman and Chapman 1980
<i>Dictyota apiculata</i>	Hawaii	-	Food	Zaneveld 1955
<i>Durvillea antarctica</i>	Chile	<i>cocbayuyo</i>	Stew	Ohni 1968
	Peru	<i>cocbayuyo</i>	Food	Masuda 1981
	New Zealand (Maori)	<i>rimurapi</i>	Roasted	Brooker and Cooper 1961
<i>Ecklonia kurome</i>	China	<i>kunbu,</i> <i>miangichai</i>	Food	Tseng 1983
<i>Ecklonia stolonifera</i>	Japan	<i>kizame arame</i>	Soup, sauce, stew	Chapman and Chapman 1980
<i>Eisenia bicyclis</i>	Japan	<i>aramae</i>	Soup, stew	Subba Rao 1965
<i>Endoracne bingbamiae</i>	China	-	Food	Tseng 1983
<i>Fucus</i> sp.	Alaska (Chugach)	-	Food	Ager and Ager 1980
	Greenland (Inuit)	<i>mikarkat</i>	Food	Hoygaard 1937
<i>Heterocordaria abietina</i>	Japan	<i>matsumo</i>	Sauce	Subba Rao 1965
<i>Hizikia fusiforme</i>	Japan	<i>bijiki</i>	Soup, stew, salad	Subba Rao 1965
<i>Hydroclatbrus clatbratus</i>	Philippines	<i>balbalulang</i>	Salad	Velasquez 1972
<i>Isbiga okamurai</i>	China	<i>teding cai</i>	Food	Tseng 1983
<i>Isbiga sinicola</i>	China	<i>bai dai</i>	Food	Tseng 1983
<i>Kjellmaniella gyrate</i>	Japan	-	Soup	Chapman and Chapman 1980
<i>Laminaria</i> sp.	New Zealand	<i>rimu roa</i>	Roasted	Goldie 1904
<i>Laminaria angustata</i>	Japan	<i>kizami-</i> <i>kombu</i>	Soup	Chapman and Chapman 1980
		<i>chimi-kombu</i>	Vegetable	Chapman and Chapman 1980
<i>Laminaria cichorioides</i>	Japan		Food	Chapman and Chapman 1980
<i>Laminaria diabolica</i>	Japan	<i>kuro-tororo</i> <i>kombu</i>	Food	Chapman and Chapman 1980
<i>Laminaria digitata</i>	Scotland	-	Food	Johnston 1970

(continued)

Table II.C.1.2. (Continued)

Species name	Country	Local name	Use	Reference
Phacophyta				
<i>Laminaria japonica</i>	Japan	<i>ma-kombu</i>	Sweetmeat	Subba Rao 1965
	Japan	<i>ori-kombu</i>	Food	Subba Rao 1965
	China	<i>bai'tai</i>	Food	Simoons 1991
<i>Laminaria digitata</i>	Japan	-	Food	Johnston 1970
<i>Laminaria ochotensis</i>	Japan	<i>risbiri-kombu</i>	Food	Chapman and Chapman 1980
<i>Laminaria religiosa</i>	Japan	<i>bosome-kombu</i> , <i>saimatsu-kombu</i>	Food	Chapman and Chapman 1980
<i>Laminaria saccharina</i>	France, Great Britain, Ireland	-	Fresh	Chapman and Chapman 1980
<i>Laminaria yezoensis</i>	Japan	-	Food	Chapman and Chapman 1980
<i>Macrocystus integrifolia</i>	British Columbia	-	Food	Turner 1975
<i>Mesogloia crassa</i>	Japan	<i>futo-mozuku</i>	Food	Chapman and Chapman 1980
<i>Mesogloia decipiens</i>	Japan	<i>mozuku</i>	Food	Chapman and Chapman 1980
<i>Nemacystus decipiens</i>	China	<i>baida</i>	Food	Tseng 1983
	Japan	<i>mozuku</i>	Food	Chapman and Chapman 1980
<i>Padina australis</i>	Indonesia	<i>agar-agar</i> , <i>daun-besar</i>	Sweetmeat	Zaneveld 1951
<i>Pelvetia siliquosa</i>	China	<i>lujiao cai</i>	Food	Tseng 1983
<i>Petalonia fascia</i>	Japan	<i>bondawara</i>	Soup, sauce	Chapman and Chapman 1980
<i>Postelsia palmaeformis</i>	California (Indian)	<i>gaye</i>	Food	Goodrich et al. 1980
<i>Sargassum</i> sp.	China	<i>bai ts'ai</i>	Tea, soup	Tseng 1935
	Hawaii	<i>limu kala</i>	Food	Schönfeld-Leber 1979
	Malaysia	-	Food	Subba Rao 1965
<i>Sargassum aguifolium</i>	Indonesia, Malaysia	<i>arien wari</i>	Raw, cooked	Zaneveld 1965
<i>Sargassum fusiformis</i>	China	<i>chu-chiau ts'ai</i>	Soup, vegetable	Tseng 1983
<i>Sargassum granuliferum</i>	Amboina	<i>arien-wari</i>	Raw, cooked	Zaneveld 1955
	Indonesia	-	Raw, cooked	Subba Rao 1965
	Malaysia	-	Raw, cooked	Zaneveld 1955
<i>Sargassum bemiphyllum</i>	China	-	Food	Tseng 1983
<i>Sargassum benslowianum</i>	China	-	Food	Tseng 1983
<i>Sargassum borneri</i>	China	-	Food	Tseng 1983
<i>Sargassum pallidum</i>	China	-	Food	Tseng 1983
<i>Sargassum polycystum</i>	Amboina	<i>arien barulu</i>	Raw	Zaneveld 1955
	Moluccas	<i>agar-agar' kupean</i>		
<i>Sargassum siliquosum</i>	Malaysia	-	Raw, cooked, pickled	Zaneveld 1955
	Philippines	<i>aragan</i>	Raw, cooked	Zaneveld 1955
<i>Scytosipbon lomentaria</i>	China	-	Food	Tseng 1983
<i>Turbinaria conoides</i>	Celebes	<i>labi-labi</i>	Salad	Zaneveld 1951
	Malaysia		Pickle	Zaneveld 1951
<i>Turbinaria ornata</i>	Malaysia	<i>agar-agar-ksong</i>	Pickle	Zaneveld 1955
	Moluccas	<i>arien essong</i>	Raw, cooked	Zaneveld 1955
<i>Undaria peterseneaniana</i>	Japan	<i>wakame</i>	-	Zaneveld 1955
<i>Undaria pinnatifida</i>	China	-	Food	Simoons 1991
	Japan	<i>wakame</i>	Food	Subba Rao 1965
<i>Undaria undarioides</i>	Japan	<i>wakame</i>	Food	Chapman and Chapman 1980
Rhodophyta				
<i>Agardhiella</i> sp.	Philippines	<i>gulaman</i>	Sweetmeat	Zaneveld 1955
<i>Acantbopeltis japonica</i>	Japan	<i>yulkitri</i>	Food	Subba Rao 1965
	Indonesia	-	Vegetable	Subba Rao 1965
<i>Acantbopora spicifera</i>	Philippines	<i>culot</i>	Salad, cooked	Velasquez 1972
<i>Abnfeltia concinna</i>	Hawaii	<i>limu akiaki</i>	Salad, baked	Schönfeld-Leber 1979
<i>Asparagopsis sanfordiana</i>	Hawaii	-	Food	Schönfeld-Leber 1979
<i>Abnfeltia taxiformis</i>	China	-	Salad, stew	Montagne 1946-7
	Hawaii	<i>limu kobu</i>	Food	Abbott 1987
<i>Bangia fusco-purpurea</i>	China	<i>bangmaocai</i>	Salad	Montagne 1946-7
	Hawaii	-	Soup	Xia and Abbott 1987
<i>Caloglossa adnata</i>	Burma	-	Raw, boiled	Zaneveld 1955
<i>Caloglossa leprieurii</i>	Burma	-	Food	Subba Rao 1965
<i>Campylaephora bypnaeides</i>	Japan	<i>yego-nori</i>	Food	Subba Rao 1965
<i>Carpopeltis flabellata</i>	Japan	<i>kome-nori</i>	Garnish	Subba Rao 1965
<i>Catenella impudica</i>	Burma	-	Salad	Zaneveld 1955
<i>Catenella nipae</i>	Burma	-	Raw, cooked	Boergesen 1938
<i>Chondria tenuissima</i>	Hawaii	<i>limu oolu</i>	Food	Schönfeld-Leber 1979
<i>Chondrus elatus</i>	China	-	Food	Johnston 1966
<i>Chondrus ocellatus</i>	China	-	Food	Tseng 1983
	Japan	<i>makuri-nori</i>	Food	Subba Rao 1965
<i>Corallopsis salicornia</i>	Indonesia	<i>bulung-buka</i>	Vegetable, jelly	Zaneveld 1955; Subba Rao 1965

Table II.C.1.2. (Continued)

Species name	Country	Local name	Use	Reference
Rhodophyta				
<i>Dermoneia frapperi</i>	China	-	Food	Lee 1965
<i>Dermoneia vulpinata</i>	China	-	Food	Tseng 1983
<i>Digenea simplex</i>	Japan	<i>makuri-nori</i>	Food	Subba Rao 1965
<i>Eucheuma edule</i>	China	<i>bat-ts'ai mu</i>	Food	Tseng 1983
	Indonesia	<i>agar-agar-besar</i>	Jelly	Zaneveld 1955
<i>Eucheuma gelatinae</i>	China	-	Food	Tseng 1983
<i>Eucheuma borridum</i>	Malaysia	-	Jelly	Zaneveld 1955
<i>Eucheuma muricatum</i>	Indonesia	-	Jelly	Subba Rao 1965
	Malaysia	-	Agar	Zaneveld 1955
	Philippines	<i>canot-canot</i>	Salad, cooked vegetable	Velasquez 1972
<i>Eucheuma serra</i>	Bali	<i>bulung</i>	Agar,	
		<i>djukut lelipan</i>	vegetable	Zaneveld 1955
<i>Eucheuma speciosa</i>	Tasmania	-	Jelly	Irving 1957
<i>Gelidiella acerosa</i>	Philippines	<i>culot</i>	Salad, cooked vegetable	Velasquez 1972
<i>Gelidium</i> sp.	China	<i>sbib-bua-tsb</i>	Agar	Tseng 1933
	Hawaii	<i>limu loloa</i>	Food	Schöfeld-Leber 1979
<i>Gelidium amansii</i>	Indonesia	-	Jelly	Subba Rao 1965
<i>Gelidium latifolium</i>	Hawaii	-	Agar	Zaneveld 1955
	Indonesia	<i>limu loloa</i>	Jelly	Subba Rao 1965
	Java	-	Agar	Zaneveld 1955
<i>Gelidium rigidum</i>	Indonesia	-	Jelly	Subba Rao 1965
	Malaysia	-	Agar	Zaneveld 1955
<i>Gigartina</i> sp.	Malaysia	-	Food	Subba Rao 1965
	New Zealand	<i>rebba</i>	Food	Schöfeld-Leber 1979
<i>Gigartina chamissoi</i>	Peru	<i>cocbayuyo</i>	Soup, stew	Polo 1977
<i>Gigartina glomerata</i>	Peru	<i>cocbayuyo</i>	Soup, stew	Polo 1977
<i>Gigartina intermedia</i>	China	-	Food	Lee 1965
<i>Gigartina teedii</i>	Japan	<i>cata-nori, sbikin-nori</i>	Food	Subba Rao 1965
<i>Gloiopeltis</i> sp.	China	-	Soup	Tseng 1933
<i>Gloiopeltis coliformis</i>	China	-	Food	Chapman and Chapman 1980
<i>Gloiopeltis furcata</i>	China	-	Food	Chapman and Chapman 1980
	Japan	<i>funori</i>	Food	Chapman and Chapman 1980
<i>Gloiopeltis tenax</i>	Taiwan	<i>funori</i>	Raw, fried	Chapman and Chapman 1980
<i>Gracilaria</i> sp.	New Zealand	<i>karengo</i>	Food	Schöfeld-Leber 1979
<i>Gracilaria confertifera</i>	China	<i>bai-mein-san</i>	Food	Tseng 1933
	Philippines	<i>gulaman</i>	Salad, cooked vegetable	Wester 1925
<i>Gracilaria coronopifolia</i>	Hawaii	<i>limu mabauea</i>	Food	Schöfeld-Leber 1979
	Philippines	<i>caocooyan</i>	Dessert	Velasquez 1972
<i>Gracilaria crassa</i>	Philippines	<i>susuldod-baybay</i>	-	Zaneveld 1955
<i>Gracilaria eucheumoides</i>	Philippines	<i>cavot-cavot</i>	Raw, cooked	Zaneveld 1955
<i>Gracilaria lichenoidea</i>	Amboina	<i>atjar</i>	Pickled	Zaneveld 1955
	Ceylon	<i>chan, chow-parsi</i>	Pudding	Subba Rao 1965
	Hawaii	<i>limu manaua</i>	Soup, jelly	Zaneveld 1955
	India	<i>conji-parsi</i>	Food	Chapman and Chapman 1980
	Malaysia	-	Pickled	Subba and Rao 1965
<i>Gracilaria salicornia</i>	Philippines	<i>susuldod-baybay</i>	Food	Velasquez 1972
<i>Gracilaria taenioides</i>	Malaysia	-	Food	Subba Rao 1965
<i>Gracilaria verrucosa</i>	Philippines	<i>susuldod-baybay</i>	Food	Galutira and Velasquez 1963
<i>Grateloupia affinis</i>	Japan	<i>kome-nori</i>	Food	Chapman and Chapman 1980
<i>Grateloupia doryphora</i>	Peru	<i>cocbayuyo</i>	Food	Polo 1977
<i>Grateloupia filicina</i>	China	-	Soup	Xia and Abbott 1987
	Japan	<i>mukade-nori</i>	Food	Subba Rao 1965
	Philippines	-	Jelly	Zaneveld 1955
<i>Grateloupia ligulata</i>	China	<i>bat-ts'ai</i>	Food	Tseng 1935
<i>Griffithsia</i> sp.	Hawaii	<i>limu moopuna-kana lipoa</i>	Food	Schöfeld-Leber 1979
<i>Gymnogongrus disciplinalis</i>	Hawaii	<i>limu vavaloli</i>	Food	Subba Rao 1965
<i>Gymnogongrus flabelliformis</i>	Japan	<i>okitsu-nori</i>	Food	Subba Rao 1965
<i>Gymnogongrus vermicularis</i>	Hawaii	<i>limu vavaloli</i>	Food	Schöfeld-Leber 1979
<i>Halymenia durvillae</i>	Hawaii	<i>limu lepeabina</i>	Food	Zaneveld 1955
	Philippines	<i>gayong-gayong</i>	Salad, cooked vegetable	Velasquez 1972
<i>Hypnea</i>	Indonesia	-	Dessert	Subba Rao 1965
<i>Hypnea armata</i>	Hawaii	<i>limu buna</i>	Food	Schöfeld-Leber 1979

(continued)

Table II.C.1.2. (Continued)

Species name	Country	Local name	Use	Reference
Rhodophyta				
<i>Hypnea cenomyce</i>	Indonesia	-	Food	Zaneveld 1955
<i>Hypnea cernicornis</i>	China	<i>sa ts'ai</i>	Food	Tseng 1935
<i>Hypnea charoides</i>	Philippines	<i>culot tipusa</i>	Salad	Velasquez 1972
<i>Hypnea divaricata</i>	Amboina	<i>arien</i>	Agar, food	Zaneveld 1955
<i>Hypnea nidifica</i>	Hawaii	<i>limu buna</i>	Food	Schönfeld-Leber 1979
<i>Hypnea musciformis</i>	China	<i>su-wei-tung</i>	Stew, jelly	Tseng 1935
<i>Iridea edulis</i>	Iceland	-	Food	Chapman and Chapman 1980
	Scotland	<i>dulse</i>	Food	Chapman and Chapman 1980
<i>Laurencia</i> sp.	Hawaii	<i>limu lipeepee</i>	Salad	Schönfeld-Leber 1979
	Polynesia	<i>lum, (limu, rimu)</i>	Cooked	Massal and Barrau 1956
<i>Laurencia botryoides</i>	Hawaii	<i>tartaripitip</i>	Raw, cooked	Zaneveld 1955
<i>Laurencia okamurai</i>	Philippines	<i>culot</i>	Salad, cooked vegetable	Velasquez 1972
<i>Laurencia papillosa</i>	Hawaii	<i>limu lipeepee</i>	Cooked salad,	Zaneveld 1955
	Philippines	<i>culot</i>	cooked vegetable	Velasquez 1972
<i>Laurencia pinnatifida</i>	Scotland, Western Europe, United States	<i>pepper dulse</i>	Seasoning	Chapman and Chapman 1980
<i>Lemanea mamilliosa</i>	India	<i>nungbam</i>	Food	Khan 1973
<i>Liagora decussata</i>	Hawaii	<i>limu puak</i>	Food	Schönfeld-Leber 1979
<i>Liagora farinosa</i>	Philippines	<i>baris-baris</i>	Food	Zaneveld 1955
<i>Macrocystis integrifolia</i>	British Columbia	giant kelp	Food	Turner 1975
<i>Nemalion belmintoides</i>	Italy, Japan	sea noodles	Food	Chapman and Chapman 1980
<i>Nemalion multifidum</i>	Japan	<i>tsukomo-nori</i>	Food	Chapman and Chapman 1980
<i>Nemalion vermiculare</i>	Japan	<i>umu-somen</i>	Food	Chapman and Chapman 1980
<i>Porphyra</i> sp.	British Columbia	-	Food	Turner and Bell 1973
	New Zealand	<i>karengo</i>	Food	Schönfeld-Leber 1979
<i>Porphyra atropurpurea</i>	Hawaii	<i>limu luau</i>	Condiment	Zaneveld 1955
	Philippines	<i>gamet</i>	Soup	
<i>Porphyra columbina</i>	Chile	<i>luche</i>	Stew	Ohni 1968
	Peru	<i>cochayuyo</i>	Stew	Polo 1977
<i>Porphyra crispata</i>	China	<i>tsu ts'ai</i>	Condiment, vegetable	Galutira and Velasquez 1963
<i>Porphyra dentata</i>	China	-	Food	Xia and Abbott 1980
<i>Porphyra laciniata</i>	California	-	Baked, raw	Yanovsky 1936
	Hebrides	-	Food	Johnston 1970
<i>Porphyra leucosticta</i>	Hawaii	<i>limu luau</i>	Food	Schönfeld-Leber 1979
	Peru	<i>cochayuyo</i>	Stew	Polo 1977
<i>Porphyra marginata</i>	China	-	Food	Xia and Abbott 1987
<i>Porphyra perforata</i>	British Columbia	-	Food	Turner 1975
<i>Porphyra suborbiculata</i>	China	<i>tsu-ts'ai</i>	Vegetable	Simoons 1991
<i>Porphyra tenera</i>	Japan	<i>awanori</i>	Food	Subba Rao 1965
<i>Porphyra vulgaris</i>	California	-	Baked, raw	Yanovsky 1936
<i>Rhodomenia palmata</i>	Canada	<i>sol</i>	Nibbled with beer	Chapman and Chapman 1980
	Iceland	<i>sol</i>	Cooked, baked	Chapman and Chapman 1980
	Ireland	-	Food	Chapman and Chapman 1980
<i>Rhodoglossum denticulatum</i>	Peru	-	Food	Polo 1977
<i>Sarcodia</i> sp.	Japan	<i>bosaka-nori, atsuba-nori</i>	Food	Subba Rao 1965
<i>Sarcodia montagneana</i>	Molluccas	-	Food	Zaneveld 1955
<i>Subria vittata</i>	South Africa	-	Jelly	Chapman and Chapman 1980

Seaweeds as Fertilizer and Animal Fodder

Seaweeds have been exploited for fertilizer by coastal farmers for centuries (if not millennia) in Europe, North America, the Mediterranean, and Asia (Booth 1965; Waaland 1981). Roman writings from the second century A.D. contain the oldest known evidence of seaweed as a fertilizer (Newton 1951). Seaweed was plowed into fields where it rotted, replenishing the soil with essential minerals. Alternatively, the seaweed was dried and burned, and the ash used as fertilizer, or the fields were "limed" with coralline algae and the sands derived from them (Waaland 1981). Regardless of the method, nitrogen, phosphate, and potassium

were delivered to the soil, along with other compounds that may serve as plant growth stimulants, by the use of algae as a fertilizer.

The Gross Chemical Composition of Algae

The gross chemical composition of algae is shown in Table II.C.1.3, where several properties are readily apparent. Although they are relatively poor in carbohydrates (including fiber) and relatively rich in lipids, microalgae are remarkably rich in protein and, thus, a good source of nutrients for humans as well as domestic animals (Aaronson, Berner, and Dubinsky 1980).

Table II.C.1.3. *The gross chemical composition of edible algae (percentage of dry weight)*

Species	Protein	Total carbohydrate plus fiber	Lipids	Total nucleic acids	Ash	HO	Reference no.
Cyanophyta							
<i>Agmenellum quadruplicatum</i>	36	32	13	-	11	-	16
<i>Nostoc commune</i>	21	60	1	-	8	11	14
<i>Nostoc phylloderma</i>	25	59	1	-	12	-	3
<i>Pbormidium tenue</i>	11	32	1	-	46	9	3
<i>Spirulina</i> sp.	64-70	-	5-7	4	-	-	13
<i>Spirulina maxima</i>	56-62	16-18	2-3	-	-	-	1
<i>Spirulina maxima</i>	60-71	13-17	6-7	3-5	6-9	-	6
<i>Spirulina platensis</i>	46-50	8-14	4-9	2-5	-	-	8
<i>Spirulina platensis</i>	63	9	3	4	10	9	5
<i>Synechococcus</i> sp.	63	15	11	5	-	-	7
Chlorophyta							
<i>Chlorella vulgaris</i>	52	25	6	-	14	5	4
<i>Chlorella vulgaris</i>	57	32	6	-	8	-	16
<i>Coelastrum proboscideum</i>	49	24	9	-	13	6	4
<i>Enteromorpha</i> sp.	20	58	0.3	-	15	-	9
<i>Enteromorpha compressa</i>	12	-	-	-	10	14	15
<i>Enteromorpha linza</i>	19	-	-	-	19	14	15
<i>Monostroma</i> sp.	20	5	1	-	15	-	9
<i>Ulva</i> sp.	26	46	1	-	23	-	9
<i>Ulva</i> sp.	15	51	-	-	16	19	15
Phaeophyta							
<i>Arthrothamnus bifidus</i>	6	52	0.7	-	17	24	15
<i>Ascophyllum nodosum</i>	5-10	42-59	2-4	-	17-20	12-15	18
<i>Hizikia fusiformis</i>	6	102	1.7	-	19	12	2
<i>Hizikia fusiformis</i>	10	57	0.5	-	-	16	15
<i>Kjellmaniella crassifolia</i>	9	62	0.6	-	28	-	9
<i>Laminaria</i> sp.	2	11	0.6	-	7	-	10
<i>Laminaria</i> sp.	6	49	1	-	21	24	15
<i>Laminaria angustata</i>	9	65	1.7	-	24	-	9
<i>Laminaria angustata</i>	9	66	2.2	-	19	-	9
<i>Laminaria japonica</i>	9	68	1.3	-	22	-	9
<i>Laminaria japonica</i>	9	66	2.2	-	23	-	9
<i>Laminaria japonica</i>	4	88	3	-	18	10	2
<i>Laminaria religiosa</i>	8	67	0.5	-	25	-	9
<i>Sargassum</i> sp.	5	35	1.3	-	25	33	19
<i>Undaria</i> sp.	3	10	0.6	-	7	-	12
<i>Undaria pinnatifida</i>	12	38	0.3	-	31	19	15
<i>Undaria pinnatifida</i>	21	8	1.7	-	31	19	15
Rhodophyta							
<i>Gelidium</i> sp.	13	68	-	-	4	-	15
<i>Gracilaria</i> sp.	4	28	-	-	4	-	15
<i>Gracilaria coronopifolia</i>	8	61	0.1	-	18	13	15
<i>Laurencia</i> sp.	9	62	1	-	19	9	15
<i>Palmaria</i> sp.	20	60	1	-	13	7	17
<i>Porphyra</i> sp.	44	46	2	-	8	-	9
<i>Porphyra laciniata</i>	29	41	2	-	19	9	17
<i>Porphyra tenera</i>	29-36	39-41	0.6	-	-	11-13	11
<i>Porphyra tenera</i>	46	64	0.5	-	12	9	2
<i>Porphyra tenera</i>	28	40	0.8	-	10	17	15
<i>Rhodomenia palmata</i>	8-35	38-74	0.2-4	-	12-37	-	19

Reference Numbers:

1. Clement, Giddey, and Merzi (1967)
2. Kishi et al. (1982)
3. Namikawa (1906)
4. El-Fouly et al. (1985)
5. Becker and Venkataraman (1984)
6. Durand-Chastel (1980)
7. Trubachev et al. (1976)
8. Tipnis and Pratt (1960)
9. Nisizawa et al. (1987)
10. Venkataraman, Becker, and Shamala (1977)
11. Arasaki and Arasaki (1983)
12. Druhl (1988)
13. Clement (1975)
14. Subbulakshmi, Becker, and Venkataraman (1976)
15. Subba Rao (1965)
16. Parsons, Stephens, and Strickland (1961)
17. Drury (1985)
18. Jensen (1972)
19. Morgan, Wright, and Simpsom (1980)

Table II.C.1.4. Amino acid content of edible algae

Species	Method of analysis	Ala	Arg	Asp	Cys	Gly	Glu
FAO standard (essential)	1	-	-	-	-	-	-
Prokaryota							
Cyanophyta							
<i>Anabaena cylindrica</i> ^a	1	9.5	6.6	13.9	tr.	6.2	14.0
<i>Calothrix</i> sp. ^a	1	8.7	6.3	12.9	tr.	5.8	14.0
<i>Nostoc commune</i> ^a	1	8.8	7.7	13.2	0.4	6.1	13.8
<i>Spirulina maxima</i> ^a	2	5.0	4.5	6.0	0.6	3.2	8.3
<i>Spirulina platensis</i> ^a	1	8.7	7.0	12.5	0.4	5.8	18.7
<i>Tolyptobrix tenuis</i> ^a	1	9.5	8.6	14.2	tr.	5.8	11.8
Eukaryota							
Chlorophyta							
<i>Chlorella pyrenoidosa</i> ^a	1	6.8	5.4	6.6	-	5.5	9.0
<i>Chlorella stigmatophora</i> ^a	1	7.9	8.6	6.5	1.4	5.4	8.5
<i>Chlorella vulgaris</i> ^a	1	9.4	7.2	10.3	1.0	6.8	12.8
<i>Dunaliella teriolecta</i> ^a	1	7.5	7.2	7.5	0.7	5.5	9.5
<i>Ulva lactuca</i> ^b	2	8.2	10.1	7.8	tr.	7.2	6.6
Phaeophyta							
<i>Ascophyllum nodosum</i> ^b	2	5.3	8.0	6.9	tr.	5.0	10.0
<i>Fucus vesiculosus</i> ^b	2	5.4	8.2	9.0	tr.	5.4	11.0
<i>Hizikia fusiforme</i> ^b	1	6.4	5.0	9.9	1.3	5.8	11.8
<i>Undaria pinnatifida</i> ^b	1	4.5	3.0	5.9	0.9	3.7	6.6
Rhodophyta							
<i>Chondrus crispus</i>	2	3.6	28.0	3.7	tr.	3.1	3.4
<i>Porphyra</i> sp. ^b	1	9.9	5.9	8.5	-	6.9	9.3
<i>Rhodomenia palmata</i> ^b	2	7.9	10.8	7.2	tr.	7.5	6.2

^aMicroalgae; ^bmacroalgae.

1 = grams/16 grams of nitrogen; 2 = percentage amino acid nitrogen as percentage of total nitrogen.

References:

1. Paoletti et al. (1973)
2. Lubitz (1961)
3. Smith and Young (1955)
4. Nisizawa et al. (1987)
5. FAO (1970)

Algae, especially marine algae, may also contain unusual amino acids not normally found in protein and iodoamino acids (Fattorusso and Piattelli 1980). By contrast, macroalgae are poor in proteins and lipids, but relatively rich in carbohydrates and minerals.

The caloric value of edible algae ranges from 4,405 to 5,410 calories per gram for the Cyanophyta; 4,700 to 4,940 for the Chlorophyta; 4,160 to 5,160 for the Phaeophyta; and 3,290 to 5,400 for the Rhodophyta (Cummins and Wuycheck 1971). The digestibility of seaweeds ranges from 39 to 73 percent, with a net energy availability of 48 to 67 percent (Kishi et al. 1982).

Amino Acid Composition

The amino acid composition of algae and their proteins is shown in Table II.C.1.4. The algae have a full complement of the amino acids found in the animal and plant proteins consumed by humans. The concen-

tration of essential amino acids required by humans is very close to the standards set for human foods by the Food and Agriculture Organization of the United Nations (FAO 1970).

Polysaccharides

Algae contain a variety of carbohydrates that serve as energy storage sources or provide structural strength to cell walls and matrices. The storage polysaccharides include mannitol and laminaran or chrysolaminaran in Chrysophyta and Phaeophyta; amylose and amylopectin in Chlorophyta; and Floridean starch in Rhodophyta. The structural polysaccharides include cellulose in the Chlorophyta (or its xylan equivalent in some Chlorophyta and Rhodophyta) and in the mannan of the Phaeophyta. Phaeophyta and Chlorophyta also contain anionic polysaccharides, such as alginic acid, in their cell walls and sulfated glucuronoxylifucans,

His	Ile	Leu	Lys	Met	Phe	Pro	Ser	Thr	Trp	Tyr	Val	Reference
-	4.0	7.0	5.5	3.5	6.0	-	-	4.0	-	-	5.0	-
1.8	6.0	10.4	5.2	2.1	5.7	4.2	6.2	6.9	1.2	6.3	7.1	1
1.7	5.7	8.7	6.8	2.6	5.5	3.4	6.4	7.1	1.1	5.5	6.3	1
1.5	4.4	9.6	5.4	1.3	5.4	4.5	5.4	6.2	1.6	4.7	5.0	1
0.9	4.7	5.6	3.0	1.6	2.8	2.7	3.2	3.2	0.8	-	4.2	2
1.8	6.3	9.8	5.2	2.9	5.3	4.0	5.4	6.1	1.4	5.5	6.6	1
1.4	6.7	8.9	4.7	1.6	5.2	3.3	4.3	4.9	1.7	3.7	7.4	1
1.5	3.6	4.1	7.8	2.0	4.8	3.7	2.7	3.4	1.5	2.9	5.8	2
2.3	3.8	9.3	13.4	1.4	5.5	5.2	4.1	5.0	-	3.7	5.7	3
2.2	4.4	10.4	6.8	2.4	6.1	5.0	5.0	5.1	1.9	4.1	6.6	1
2.5	4.3	10.7	13.9	0.8	6.6	4.1	4.5	2.6	-	4.1	5.3	3
0.2	2.8	5.0	5.8	1.0	3.0	3.4	3.9	4.2	tr.	1.6	4.9	3
1.3	2.8	4.6	4.9	0.7	2.3	2.6	3.0	2.8	tr.	0.9	3.7	3
1.6	3.0	5.0	6.0	0.4	2.6	3.3	3.5	3.3	tr.	1.2	3.9	3
0.9	6.2	0.5	2.9	3.2	5.8	4.8	3.8	3.2	0.8	3.0	10.8	4
0.8	3.7	5.9	1.1	1.8	4.5	5.3	3.2	1.1	1.8	3.7	7.8	4
0.5	2.9	8.5	3.7	2.1	3.7	3.0	2.6	5.4	1.2	1.6	6.9	4
1.1	1.7	2.6	3.3	0.5	1.3	2.1	2.2	2.0	tr.	1.0	2.6	3
1.2	4.0	7.7	2.6	3.4	5.3	4.6	4.8	3.2	1.1	2.4	9.3	4
1.1	3.1	4.6	6.4	0.6	3.1	3.6	4.8	3.9	tr.	1.6	5.1	3

fucoidan, and ascophyllan as storage polysaccharides. Additionally, the Rhodophyta contain galactans (agar and carrageenans) (see Lewin 1974; McCandless 1981). The cyanobacteria have peptidoglycan cell walls and may contain polyglucan granules or polyphosphates as energy storage molecules (Dawes 1991).

Hydrocolloids

Hydrocolloids are water-soluble gums (polysaccharides) commonly obtained from seaweed or other plants that have been employed since antiquity to thicken foods and today are used to thicken, to emulsify, or to gel aqueous solutions in many industries (Spalding 1985). The major hydrocolloids are: (1) agar, which is obtained from species of the red algal genera - *Gelidium*, *Gracilaria*, *Pterocladia*; (2) alginates, obtained from species of brown algal genera - *Ascophyllum*, *Ecklonia*, *Eisenia*, *Laminaria*, *Macrocystis*, *Nereocystis*, *Sargassum*; and (3) carrageenans, also obtained from species of the red algal genera - *Abnfeltia*, *Chondrus*, *Eucbema*, *Furcellaria*, *Gigartina*, *Gymnogongrus*, *Hypnea*, *Iridaea*, *Phyllophora*. Hydrocolloids have great impor-

tance in the food industries, where they are employed in the production of such varied products as glazes, icings, frostings, toppings, frozen foods, cereals, bread, salad dressings, flavors, sausage casings, puddings, desserts, candies, marshmallows, processed meat products, cheese, jams, pie fillings, and sauces (Spalding 1985).

Vitamins and Other Growth Factors

Algae can also be an excellent source of water-soluble and fat-soluble vitamins (Table II.C.1.5). The concentration of specific vitamins in algae varies from species to species and depends on conditions of algal growth, handling, storage, and methods of preparation for eating, as well as on the number of microorganisms found on the surface of macroalgae, which also may be responsible for some of the B vitamins attributed to macroalgae (Kong and Chan 1979).

Tocopherols are a metabolic source of vitamin E, and they are found in most types of algae as alphatocopherol. The Fucaceae family of brown algae contains delta-homologues of tocopherol as well as alphatocopherol (Jensen 1969), and seaweeds contain 7 to 650 micrograms per gram (Ragan 1981).

Table II.C.1.5. Vitamin content of edible algae

Species	A (IU/ 100 g)	D (µg/g)	E (µg/g)	Thiamine (µg/g)	Riboflavin (µg/g)	B ₆ (µg/g)	Nicotinate (µg/g)
Prokaryota							
Cyanophyta							
<i>Anabaena cylindrica</i> ^a	-	-	4,000	-	55,000	7,000	78
<i>Spirulina</i> sp. ^a	23,000	-	-	37	46	7	-
<i>Spirulina platensis</i> ^a	-	-	-	28	33	1	-
Eukaryota							
Chlorophyta							
<i>Caulerpa racemosa</i> ^b	-	-	-	1	2	-	21
<i>Chlamydomonas reinhardtii</i> ^a	-	-	4	-	-	-	-
<i>Enteromorpha</i> sp. ^b	500	-	-	0.4	5	-	10
<i>Enteromorpha linza</i> ^b	2,900	3	-	2	1	-	28
<i>Monostroma nitidum</i> ^b	-	-	-	1	9	-	10
<i>Ulva</i> sp. ^b	960	-	0.6	0.3	-	-	80
<i>Ulva pertusa</i> ^b	-	-	0.9	3	-	-	8
Phaeophyta							
<i>Ascophyllum nodosum</i> ^b	-	-	-	1-5	5-10	-	10-30
<i>Colpomenia sinuosa</i> ^b	-	-	-	0.3	5	-	5
<i>Dictyota dichotoma</i> ^b	-	-	-	0.8	6	-	15
<i>Dictyopteris prolifera</i> ^a	-	-	-	0.5	4	-	18
<i>Ecklonia cava</i> ^b	-	-	-	1	3	-	19
<i>Eisenia bicyclis</i> ^b	-	-	-	0.2	0.2	-	26
<i>Hizikia fusiforme</i> ^b	450	16	-	0.3	3	-	7
<i>Hydroclathrus clathratus</i> ^b	-	-	-	0.3	3	-	4
<i>Laminaria</i> sp. ^b	440	-	-	0.9	0.2	0.3	30
<i>Padina arboriscens</i> ^b	-	-	-	0.3	1	-	14
<i>Sargassum fulvellum</i> ^b	-	-	-	0.4	5	-	9
<i>Sargassum nigrifolium</i> ^b	-	-	-	0.4	6	-	17
<i>Sargassum thunbergii</i> ^b	-	-	-	0.4	5	-	5
<i>Spathoglossum pacificum</i> ^b	-	-	-	0.4	0.8	-	25
<i>Undaria pinnatifida</i> ^b	140	-	-	1	1	-	100
Rhodophyta							
<i>Chondrococcus japonicus</i> ^b	-	-	-	1	11	-	8
<i>Chondrus ocellatus</i> ^b	-	-	-	2	15	-	30
<i>Gelidium amansii</i> ^b	-	-	-	7	18	-	20
<i>Gloiopeltis tenax</i> ^b	-	-	-	3	15	-	24
<i>Gracilaria gigas</i> ^b	800	-	-	2	1	-	8
<i>Gracilaria textorii</i> ^b	-	-	-	5	7	-	34
<i>Grateloupia ramosissima</i> ^b	-	-	-	1	6	-	25
<i>Hypnea charoides</i> ^b	-	-	-	1	3	-	22
<i>Laurencia okamura</i> ^b	-	-	-	0.5	10	-	39
<i>Lomentaria catenata</i> ^b	-	-	-	1	3	-	24
<i>Palmaria</i> sp. ^b	-	2	-	0.7	10	-	69
<i>Porphyra laciniata</i> ^b	-	472	-	-	-	-	115
<i>Porphyra tenera</i> ^b	-	44,500	-	2	23	10	68
<i>Porphyra tenera</i> ^b	-	44,500	-	3	12	-	100

^aMicroalgae; ^bmacroalgae

References:

1. Kanazawa (1963)
2. Arasaki and Arasaki (1983)
3. Jensen (1972)
4. Drury (1985)
5. Aaronson et al. (1977)
6. Becker and Venkataraman (1984)
7. Jassley (1988)

Pantothenate (µg/g)	Folate (µg/g)	Biotin (µg/g)	Lipoate (µg/g)	B ₁₂ (µg/g)	Choline (µg/g)	Inositol (µg/g)	Ascorbate (mg/100 g)	Reference
88,000	15,000	-	-	-	-	-	2,000	5
5	510	46	-	1,700	-	-	20	7
-	-	-	-	-	-	-	-	6
6	612	131	295	149	-	581	-	1
-	9,000	260	-	-	-	-	-	5
-	429	-	-	13	-	-	-	2
-	270	198	175	98	358	581	-	1
4	429	115	575	13	79	219	75-80	1
-	118	-	-	63	-	-	-	2
2	118	224	420	63	61	330	27-41	1
-	200-1,000	100-400	-	4	-	-	-	3
3	46	136	540	77	406	146	-	1
0.7	521	187	500	10	77	125	-	1
4	170	163	485	17	242	151	-	1
0.5	-	209	90	3	27	690	-	1
-	-	-	-	-	-	-	-	2
2	218	237	230	6	262	379	-	1
3	857	181	330	66	33	328	0-92	1
-	-	-	-	3	49	405	-	1
2	542	160	230	4	618	1,131	-	1
9	308	282	270	47	95	60	-	1
0.5	249	159	410	21	24	197	-	1
9	308	282	270	47	28	566	-	1
0.3	566	150	655	7	87	144	-	1
-	-	-	-	-	-	-	15	2
1	97	40	250	220	1,337	449	-	1
7	-	69	700	89	856	111	16	1
1	782	61	570	36	4,885	443	-	1
6	676	37	330	15	319	163	-	1
2	304	18	495	212	1,492	324	-	1
10	668	153	985	76	230	668	-	1
2	719	82	530	29	1,119	55	-	1
7	540	95	355	27	636	257	-	1
9	763	95	300	100	1,346	89	4	1
12	220	90	625	25	240	263	-	1
-	-	-	-	-	-	-	5	4
-	-	60	-	-	-	-	17	4
-	88	294	790	290	2,920	62	10-831	1
-	-	-	-	-	-	-	20	2

Algae can also contain vitamin C and beta-carotene, which are among the nutrients presently thought to protect cells against powerful oxidizing agents such as ozone, lipid peroxides, and nitrogen dioxide, and are, consequently, recommended in the diet (Calabrese and Horton 1985; Kennedy and Liebler 1992; Krinsky 1992). Vitamin C is found in all seaweeds in concentra-

tions up to 10 milligrams per gram (Ragan 1981). The fat-soluble vitamin A is synthesized from beta-carotene by humans, and beta-carotene is found in comparatively large amounts in many algae eaten by humans, such as those species in the Chlorophyta, Phaeophyta, Rhodophyta taxa, and blue-green bacteria, as well as other algal taxa (Goodwin 1974).

Table II.C.1.6. *The range of fatty acids found in edible algae*

Fatty acid carbon no.	Percentage range of total fatty acid			
	Cyanophyta 29 ^a	Chlorophyta 28 ^a	Phaeophyta 11 ^a	Rhodophyta 11 ^a
14:0	-	1-12	10-12	1-10
16:0	9-54	14-35	15-36	18-53
16:1	4-45	1-29	2-32	2-7
16:2	1-14	1-8	-	0.1-1
16:3	-	1-12	-	-
16:4	-	3-19	-	-
18:0	1-9	1-53	1	1-11
18:1	2-26	2-46	17-19	3-34
18:2	1-37	4-34	2-9	1-21
18:3 (gamma)	5-35	1-6	-	-
18:3 (alpha)	2-18	1-34	7-8	0.4-2
18:4	-	1-29	6-7	0.5-1
20:0	-	0.5-1	-	-
20:1	-	1-17	-	-
20:2	-	1-2	1	-
20:3	-	1-2	1	1-7
20:4	-	1-4	10-11	5-36
20:5	-	2-10	8	17-24
22:0	-	1-3	-	1.5
22:5	-	2-6	-	-
22:6	-	-	-	-

Sources: Adapted from Shaw (1966), Watanabe (1970), and Wood (1974).

^aNumber of species examined for fatty acids in above references.

Ash and Water Content of Algae

As Table II.C.1.3 indicates, seaweeds mostly comprise water and ash. T. Yamamoto and colleagues (1979) examined the ash content of marine algae collected in Japanese waters and found it to vary from 4 to 76 percent. J. H. Ryther, J. A. De Boer, and B. E. Lapointe (1978) found that the wet weight of several seaweeds cultured in Florida consisted of about 10 to 16 percent dry material, 32 to 50 percent of which was minerals. The seaweed ash of *Eisenia bicyclis* includes the following elements (in order of decreasing concentration): potassium, calcium, sodium, phosphorus, magnesium, strontium, zinc, iron, boron, aluminum, copper, titanium, nickel, vanadium, chromium, cobalt, molybdenum, and gallium (Yamamoto et al. 1979). Marine algae can contain up to 5 milligrams of iodine per gram of dried algae, although the amount varies from species to species and from one part of the seaweed to another (Grimm 1952).

Lipids

Algal lipids include the saponifiable lipids: fatty acids (acylglycerols, phosphoglycerides, spingolipids, and waxes) and the nonsaponifiable lipids (terpenes, steroids, prostaglandins, and hydrocarbons). As already noted, microalgae are far richer in lipids than macroalgae (Table II.C.1.3), and algae that grow in colder waters contain more unsaturated fatty acids than do algae that thrive in warm waters. Algae supply

nutrients, especially lipids, when they are eaten directly as food, but they can also pass on their nutrients indirectly when they are consumed by zooplankton, which are subsequently eaten by other invertebrates and vertebrates. Algal nutrients are then passed along to humans when they eat invertebrates and vertebrates, such as shellfish and shrimp, and fish or fish-eating birds and mammals, respectively.

Fatty Acids

Algae contain varying amounts of saturated and unsaturated fatty acids (Table II.C.1.6). Algae are rich in alpha- and gamma-linolenic acids and unusually rich in polyunsaturated fatty acids.

Steroids

Steroids are found in all eukaryotic algae, composing between 0.02 to 0.38 percent of their dry weight. Many different steroids, including sterols, are specific to one species of algae. For example, cholesterol is found in large amounts in Rhodophyta, but Phaeophyta and blue-green bacteria contain comparatively smaller amounts (Nes 1977).

Essential Oils

Seaweeds often have a characteristic odor of iodine and bromine when freshly isolated from the sea. Some brown seaweeds, however, may have a unique odor, due to 0.1 to 0.2 percent (wet weight) of essen-

tial oils (of which the major hydrocarbons are dictyopterene A and B), which impart a characteristic odor and flavor to the seaweeds belonging to the genus *Dictyopteris* found in Hawaii and known there as *limu lipoa*. *Limu lipoa* is used in Hawaii to season raw fish, meats, and stews. It was one of the only spices known in old Hawaii and was used in Hawaiian recipes much the same as other cultures used pepper and sage (Moore 1976).

Pharmacologically Active Compounds

Algae can possess small amounts of pharmacologically active molecules that affect humans. The polyunsaturated fatty acids of marine seaweeds in the diet may reduce blood pressure in humans; eicosanoids (including prostaglandins) are important biological regulators in humans; and eicosapentaenoic acids may influence the inflammatory process in humans (Beare-Rogers 1988). Algae are known to produce relatively large amounts of polyunsaturated fatty acids (Ackman 1981) and eicosanoids (Jiang and Gerwick 1991). Small amounts of 3-iodo- and 3,5-diiodotyrosine, triiodothyronine, and thyroxine are found in many brown and red seaweeds (Ericson and Carlson 1953; Scott 1954). Betaines have been found in most marine algae and several seaweeds regularly eaten in Japan (Blunden and Gordon 1986). These include the green algae, *Monostroma-nitidum*, *Ulva pertusa*, *Enteromorpha compressa*, and *E. prolifera*, and the red alga, *Porphyra tenera*, which, in experiments, have lowered the blood cholesterol of rats (Abe 1974; Abe and Kaneda 1972, 1973, 1975).

Some freshwater and marine cyanobacteria contain protease inhibitors, which can affect protein digestion; about 7 percent of the algal cultures examined by R. J. P. Cannell and colleagues (1987) were positive for protease inhibitors.

Algae contain a large variety of phenols, especially the marine brown and red algae. Some phenols are antimicrobial, and others have been found to be anticancer agents (Higa 1981).

Anthelmintic compounds have been associated with seaweeds for hundreds of years, if not longer. More recently, anthelmintic compounds with neurotoxic properties, such as alpha-kainic acid and domoic acid, were isolated from *Digenea simplex* (Ueyanagi et al. 1957), *Chondria armata* (Daigo 1959), and also from diatoms (Fritz et al. 1992).

H. Noda and colleagues (1990) have reviewed the antitumor activity of the aqueous extracts of several seaweeds, as well as 46 species of marine algae (4 green, 21 brown, and 21 red algae). Certain species of brown algae (*Scytosiphon lomentaria*, *Lessonia nigrescens*, *Laminaria japonica*, *Sargassum ringolianism*), red algae, (*Porphyra yezoensis* and *Eucheuma gelatiniae*), and the green alga *Enteromorpha prolifera* were found to have significant activity against Ehrlich carcinoma in mice. However, the cancer-causing polyaromatic hydrocarbon, 3,4-ben-

zopyrene, has been reported in commercially sold nori (*Porphyra* spp.) (Shirotori 1972; Shiraishi, Shirotori, and Takahata 1973).

Algae, like other plants, contain a variety of compounds, such as amino acids, ascorbic acid, carotenoids, cinnamic acids, flavonoids, melanoidins, peptides, phosphatides, polyphenols, reductones, tannins, and tocopherols. These molecules may act as reducing agents or free radical interrupters, as singlet oxygen quenchers, and as inactivators of preoxidant metals, thus preventing the formation of powerful oxidizers and mutagens (Tutour 1990).

One seaweed (as yet unidentified), called *limu mualea*, was thought to be highly poisonous in Hawaii (Schönfeld-Leber 1979), and a number of algae and some cyanobacteria produce secondary metabolites that are toxic to humans. Certainly, as with other aquatic organisms, eating algae from polluted waters is hazardous because of potential contamination by microbial pathogens (viruses, bacteria, fungi, or protozoa), toxic metals or ions, pesticides, industrial wastes, or petroleum products (see Jassley 1988 for a review).

Sheldon Aaronson

This work was funded, in part, by PSC/CUNY and Ford Foundation Urban Diversity research awards.

Bibliography

- Aaronson, S. 1986. A role for algae as human food in antiquity. *Food and Foodways* 1: 311-15.
- Aaronson, S., T. Berner, and Z. Dubinsky. 1980. Microalgae as a source of chemicals and natural products. In *Algae biomass*, ed. G. Shelef and C. J. Soeder, 575-601. Amsterdam.
- Aaronson, S., S. W. Dhawale, N. J. Patni, et al. 1977. The cell content and secretion of water-soluble vitamins by several freshwater algae. *Archives of Microbiology* 112: 57-9.
- Abbott, I. A. 1978. The uses of seaweed as food in Hawaii. *Economic Botany* 32: 409-12.
- Abe, S. 1974. Occurrence of homoserine betaine in the hydrolysate of an unknown base isolated from a green alga. *Japanese Fisheries* 40: 1199.
- Abe, S., and T. Kaneda. 1972. The effect of edible seaweeds on cholesterol metabolism in rats. In *Proceedings of the Seventh International Seaweed Symposium*, ed. K. Nisizawa, 562-5. Tokyo.
- 1973. Studies on the effects of marine products on cholesterol metabolism in rats. VIII. The isolation of hypocholesterolemic substance from green laver. *Bulletin of the Japanese Society of Scientific Fisheries* 39: 383-9.
- 1975. Studies on the effects of marine products on cholesterol metabolism in rats. XI. Isolation of a new betaine, ulvaline, from a green laver *Monostroma-nitidum* and its depressing effect on plasma cholesterol levels. *Bulletin of the Japanese Society of Scientific Fisheries* 41: 567-71.

- Abe, S., M. Uchiyama, and R. Sato. 1972. Isolation and identification of native auxins in marine algae. *Agricultural Biological Chemistry* 36: 2259-60.
- Ackman, R. G. 1981. Algae as sources for edible lipids. In *New sources of fats and oils*, ed. E. H. Pryde, L. H. Princen, and K. D. Malcherjee, 189-219. Champaign, Ill.
- Ager, T. A., and L. P. Ager. 1980. Ethnobotany of the Eskimos of Nelson Island [Alaska]. *Arctic Anthropology* 17: 27-49.
- Aldave-Pajaras, A. 1969. Cushuro algas azul-verdes utilizados como alimento en la región altoandina Peruana. *Boletín de la Sociedad Botánica de la Libertad* 1: 5-43.
1985. High Andean algal species as hydrobiological food resources. *Archiv für Hydrobiologie und Beibef: Ergebnisse der Limnologie* 20: 45-51.
- Arasaki, S., and T. Arasaki. 1983. *Vegetables from the sea*. Tokyo.
- Baker, J. T., and V. Murphy, eds. 1976. Compounds from marine organisms. In *Handbook of marine science: Marine products*, Vol. 3, Section B, 86. Cleveland, Ohio.
- Barrau, J. 1962. *Les plantes alimentaires de l'océanie originaire*. Marseille.
- Beare-Rogers, J. 1988. Nutritional attributes of fatty acids. *Journal of the Oil Chemists' Society* 65: 91-5.
- Becker, E. W., and L. V. Venkataraman. 1984. Production and utilization of the blue-green alga *Spirulina* in India. *Biomass* 4: 105-25.
- Birket-Smith, K. 1953. *The Chugach Eskimo*. Copenhagen.
- Biswas, K. 1953. The algae as substitute food for human and animal consumption. *Science and Culture* 19: 246-9.
- Blunden, G., and S. M. Gordon. 1986. Betaines and their sulphonic analogues in marine algae. *Progress in Phycological Research* 4: 39-80.
- Blunden, G., S. M. Gordon, and G. R. Keysell. 1982. Lysine betaine and other quaternary ammonium compounds from British species of Laminariales. *Journal of Natural Products* 45: 449-52.
- Boergesen, F. 1938. *Catenella nipae* used as food in Burma. *Journal of Botany* 76: 265-70.
- Booth, E. 1965. The manual value of seaweed. *Botanica Marina* 8: 138-43.
- Brooker, S. G., R. C. Combie, and R. C. Cooper. 1989. Economic native plants of New Zealand. *Economic Botany* 43: 79-106.
- Brooker, S. G., and R. C. Cooper. 1961. *New Zealand medicinal plants*. Auckland.
- Browman, D. L. 1980. El manejo de la tierra árida del altiplano del Perú y Bolivia. *América Indígena* 40: 143-59.
1981. Prehistoric nutrition and medicine in the Lake Titicaca basin. In *Health in the Andes*, ed. J. W. Bastien and J. M. Donahue, 103-18. Washington, D.C.
- Calabrese, E. J., and J. H. M. Horton. 1985. The effects of vitamin E on ozone and nitrogen dioxide toxicity. *World Review of Nutrition and Diet* 46: 124-47.
- Cannell, R. J. P., S. J. Kellam, A. M. Owsianka, and J. M. Walker. 1987. Microalgae and cyanobacteria as a source of glucosidase inhibitors. *Journal of General Microbiology* 133: 1701-5.
- Chapman, V. J. 1970. *Seaweeds and their uses*. London.
- Chapman, V. J., and D. J. Chapman. 1980. *Seaweeds and their uses*. London.
- Chase, F. M. 1941. Useful algae. *Smithsonian Institution, annual report of the board of regents*, 401-52.
- Chu, H.-J., and C.-K. Tseng. 1988. Research and utilization of cyanophytes in China: A report. *Archives of Hydrobiology, Supplement* 80: 573-84.
- Clement, G. 1975. Spirulina. In *Single cell protein II*, ed. S. R. Tannenbaum and D. I. C. Wang, 467-74. Cambridge, Mass.
- Clement, G., C. Giddey, and R. Merzi. 1967. Amino acid composition and nutritive value of the alga *Spirulina maxima*. *Journal of the Science of Food and Agriculture* 18: 497-501.
- Cobo, B. 1956. *Obras*, ed. P. Francisco Mateos. 2 vols. Madrid.
- Cummins, K. W., and J. C. Wuycheck. 1971. *Caloric equivalents for investigations in ecological energetics*. Internationale Vereinigung für theoretische und angewandte Limnologie. Monograph Series No. 18. Stuttgart.
- Daigo, K. 1959. Studies on the constituents of *Chondria armata* III. Constitution of domoic acid. *Journal of the Pharmaceutical Society of Japan* 79: 356-60.
- Dangeard, P. 1940. On a blue alga edible for man: *Arthrospira platensis* (Nordst.) Gomont. *Actes de la Société Linéenne de Bordeaux* 91: 39-41.
- Dawes, E. A. 1991. Storage polymers in prokaryotes. *Society of General Microbiology Symposium* 47: 81-122.
- Delpeuch, F., A. Joseph, and C. Cavelier. 1975. Consommation alimentaire et apport nutritionnel des algues bleues (*Oscillatoria platensis*) chez quelques populations du Kanem (Tchad). *Annales de la Nutrition et de l'Alimentation* 29: 497-516.
- Dillehay, T. D. 1989. *Monte Verde*. Washington, D.C.
- Druehl, L. D. 1988. Cultivated edible kelp. In *Algae and human affairs*, ed. C. A. Lembi and J. R. Waaland, 119-47. Cambridge.
- Drury, H. M. 1985. Nutrients in native foods of southeastern Alaska. *Journal of Ethnobiology* 5: 87-100.
- Durand-Chastel, H. 1980. Production and use of *Spirulina* in Mexico. In *Algae biomass*, ed. G. Shelef and C. J. Soeder, 51-64. Amsterdam.
- Eidlihtz, M. 1969. *Food and emergency food in the circum-polar area*. Uppsala, Sweden.
- Elenkin, A. A. 1931. On some edible freshwater algae. *Priroda* 20: 964-91.
- El-Fouly, M., F. E. Abdalla, F. K. El Baz, and F. H. Mohn. 1985. Experience with algae production within the Egypto-German microalgae project. *Archiv für Hydrobiologie Beibef: Ergebnisse der Limnologie* 20: 9-15.
- Ericson, L.-E., and B. Carlson. 1953. Studies on the occurrence of amino acids, niacin and pantothenic acid in marine algae. *Arkiv for Kemi* 6: 511-22.
- FAO (Food and Agriculture Organization of the United Nations). 1970. *Amino-acid content of foods and biological data on proteins*. FAO Nutritional Studies No. 24. Rome.
- Fattorusso, E., and M. Piattelli. 1980. Amino acids from marine algae. In *Marine natural products*, ed. P. J. Scheuer, 95-140. New York.
- Feldheim, W., H. D. Payer, S. Saovakntha, and P. Pongpaew. 1973. The uric acid level in human plasma during a nutrition test with microalgae in Thailand. *Southeast Asian Journal of Tropical Medicine and Public Health* 4: 413-16.
- Fritz, L., A. M. Quilliam, J. L. C. Wright, et al. 1992. An outbreak of domoic acid poisoning attributed to the pennate diatom *Pseudonitzschia australis*. *Journal of Phycology* 28: 439-42.
- Furst, P. T. 1978. Spirulina. *Human Nature* (March): 60-5.
- Gade, D. W. 1975. *Plants, man and the land in the Vilcanota valley of Peru*. The Hague, the Netherlands.
- Gaitan, E. 1990. Goitrogens in food and water. *Annual Review of Nutrition* 10: 21-39.
- Galutira, E. C., and C. T. Velasquez. 1963. Taxonomy, distribution and seasonal occurrence of edible marine algae in Ilocos Norte, Philippines. *Philippine Journal of Science* 92: 483-522.

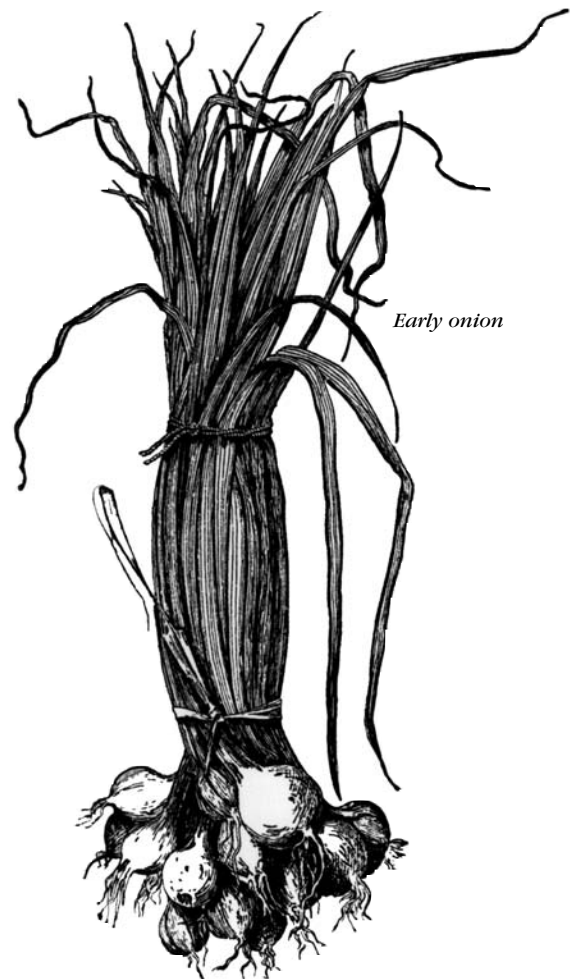
- Goldie, W. H. 1904. Maori medical lore. *Transactions of the New Zealand Institute* 37: 1-120.
- Goodrich, J., C. Lawson, and V. P. Lawson. 1980. *Kasharya pomo plants*. Los Angeles.
- Goodwin, T. W. 1974. Carotenoids and biliproteins. In *Algal physiology and biochemistry*, ed. W. D. P. Stewart, 176-205. Berkeley, Calif.
- Grimm, M. R. 1952. Iodine content of some marine algae. *Pacific Science* 6: 318-23.
- Guaman Poma de Ayala, F. 1965-6. *La nueva cronica y buen gobierno*, Vol. 3. Lima.
- Gunther, E. 1945. *Ethnobotany of western Washington*. Seattle.
- Güven, K. C., E. Guler, and A. Yucel. 1976. Vitamin B-12 content of *Gelidium capillaceum* Kütz. *Botanica Marina* 19: 395-6.
- Hasimoto, Y. 1979. *Marine toxins and other bioactive marine metabolites*. Tokyo.
- Higa, T. 1981. Phenolic substances. In *Marine natural products*, ed. P. Scheuer, 93-145. New York.
- Hiroe, M. 1969. *The plants in the Tale of Genji*. Tokyo.
- Hoygaard, A. 1937. *Skrifter om Svalbard og Ishavet*. Oslo.
- Irving, F. R. 1957. Wild and emergency foods of Australian and Tasmanian aborigines. *Oceania* 28: 113-42.
- Izawa, K., and T. Nishida. 1963. Monkeys living in the northern limits of their distribution. *Primates* 4: 67-88.
- Jao, C. 1947. *Prasiola Yunnanica* sp. nov. *Botanical Bulletin of the Chinese Academy* 1: 110.
- Jassley, A. 1988. *Spirulina*: A model algae as human food. In *Algae and human affairs*, ed. C. A. Lembi and J. R. Waaland, 149-79. Cambridge.
- Jensen, A. 1969. Tocopherol content of seaweed and seaweed meal. I. Analytical methods and distribution of tocopherols in benthic algae. *Journal of Scientific Food and Agriculture* 20: 449-53.
1972. The nutritive value of seaweed meal for domestic animals. In *Proceedings of the Seventh International Seaweed Symposium*, ed. K. Nisizawa, 7-14. New York.
- Jiang, Z. D., and W. H. Gerwick. 1991. Eicosanoids and the hydroxylated fatty acids from the marine alga *Gracilaria lemaneiformis*. *Phytochemistry* 30: 1187-90.
- Johnston, H. W. 1966. The biological and economic importance of algae. Part 2. *Tuatara* 14: 30-63.
1970. The biological and economic importance of algae. Part 3. Edible algae of fresh and brackish water. *Tuatara* 18: 17-35.
- Kanazawa, A. 1963. Vitamins in algae. *Bulletin of the Japanese Society for Scientific Fisheries* 29: 713-31.
- Kennedy, T. A., and D. C. Liebler. 1992. Peroxyl radical scavenging by B-carotene in lipid bilayers. *Journal of Biological Chemistry* 267: 4658-63.
- Khan, M. 1973. On edible Lemanea Bory de St. Vincent - a fresh water red alga from India. *Hydrobiologia* 43: 171-5.
- Kishi, K., G. Inoue, A. Yoshida, et al. 1982. Digestibility and energy availability of sea vegetables and fungi in man. *Nutrition Reports International* 26: 183-92.
- Kong, M. K., and K. Chan. 1979. Study on the bacterial flora isolated from marine algae. *Botanica Marina* 22: 83-97.
- Krinsky, N. I. 1992. Mechanism of action of biological antioxidants. *Proceedings of the Society for Experimental Biology and Medicine* 200: 248-54.
- Lagerheim, M. G. de. 1892. La "Yuyucha." *La Nuevo Notarista* 3: 137-8.
- Lee, K.-Y. 1965. Some studies on the marine algae of Hong Kong. II. Rhodophyta. *New Asia College Academic Annual* 7: 63-110.
- Léonard, J. 1966. The 1964-65 Belgian Trans-Saharan Expedition. *Nature* 209: 126-7.
- Léonard, J., and P. Compère. 1967. *Spirulina platensis*, a blue alga of great nutritive value due to its richness in protein. *Bulletin du Jardin botanique naturelle de l'État à Bruxelles (Supplement)* 37: 1-23.
- Lewin, R. A. 1974. Biochemical taxonomy. In *Algal physiology and biochemistry*, ed. W. D. Stewart, 1-39. Berkeley, Calif.
- Lewmanomont, K. 1978. Some edible algae of Thailand. Paper presented at the Sixteenth National Conference on Agriculture and Biological Sciences. Bangkok.
- Lubitz, J. A. 1961. *The protein quality, digestibility and composition of Chlorella 171105*. Research and Development Department, Chemical Engineering Section. General Dynamics Corporation Biomedical Laboratory Contract No. AF33(616)7373, Project No. 6373, Task No. 63124. Groton, Conn.
- Madlener, J. C. 1977. *The sea vegetable book*. New York.
- Massal, E., and J. Barrau. 1956. *Food plants of the South Sea Islands*. Noumea, New Caledonia.
- Masuda, S. 1981. Cochayuyo, Macha camaron y higos cargueados. In *Estudios etnográficos del Perú meridional*, ed. S. Masuda, 173-92. Tokyo.
1985. Algae. . . . In *Andean ecology and civilization*, ed. S. Masuda, I. Shimada, and C. Morris, 233-50. Tokyo.
- Matsuzawa, T. 1978. The formative site of Las Haldas, Peru: Architecture, chronology and economy. *American Antiquity* 43: 652-73.
- McCandless, E. L. 1981. Polysaccharides of seaweeds. In *The biology of seaweeds*, ed. C. S. Lobban and M. J. Wynne, 558-88. Berkeley, Calif.
- Michanek, G. 1975. *Seaweed resources of the ocean*. FAO Fisheries Technical Paper No. 138. Rome.
- Miyashita, A. 1974. *The seaweed. The cultural history of material and human being*. Tokyo.
- Montagne, M. C. 1946-7. Un dernier mot sur le *Nostoc edule* de la Chine. *Revue botanique* 2: 363-5.
- Moore, R. E. 1976. Chemotaxis and the odor of seaweed. *Lloydia* 39: 181-91.
- Morgan, K. C., J. L. C. Wright, and F. J. Simpsom. 1980. Review of chemical constituents of the red alga, *Palmaria palmata* (dulse). *Economic Botany* 34: 27-50.
- Moseley, M. E. 1975. *The maritime foundations of Andean civilization*. Menlo Park, Calif.
- Moseley, M. E., and G. R. Willey. 1973. Aspero, Peru: A reexamination of the site and its implications. *American Antiquity* 38: 452-68.
- Namikawa, S. 1906. Fresh water algae as an article of human food. *Bulletin of the College of Agriculture. Tokyo Imperial University* 7: 123-4.
- Nes, W. R. 1977. The biochemistry of plant sterols. *Advances in Lipid Research* 15: 233-324.
- Newton, L. 1951. *Seaweed utilisation*. London.
- Nisizawa, K., H. Noda, R. Kikuchi, and T. Watanabe. 1987. The main seaweed foods in Japan. *Hydrobiologia* 151/2: 5-29.
- Noda, H., H. Amano, K. Arashima, and K. Nisizawa. 1990. Antitumor activity of marine algae. *Hydrobiologia* 204/5: 577-84.
- Norton, H. H. 1981. Plant use in Kaigani Haida culture. Correction of an ethnohistorical oversight. *Economic Botany* 35: 434-49.
- Oberg, K. 1973. *The social economy of the Tlingit Indians*. Seattle, Wash.
- Ohni, H. 1968. Edible seaweeds in Chile. *Japanese Society of Physiology Bulletin* 16: 52-4.
- Ortega, M. W. 1972. Study of the edible algae of the Valley of Mexico. *Botanica Marina* 15: 162-6.

- Ostermann, H. 1938. Knud Rasmussen's posthumous notes on the life and doings of east Greenlanders in olden times. *Meddelelser Om Grønland*, 109.
- Paoletti, C., G. Florenzano, R. Materassi, and G. Caldini. 1973. Ricerche sulla composizione delle proteine di alcune ceppe coltivati di microalghe verdi e verdi-azzurre. *Scienze e Tecnologia degli alimenti* 3: 171-6.
- Parsons, M. H. 1970. Pre-ceramic subsistence on the Peruvian coast. *American Antiquity* 35: 292-304.
- Parsons, T. R., K. Stephens, and J. D. H. Strickland. 1961. On the chemical composition of eleven species of marine phytoplankton. *Journal of the Fisheries Research Board of Canada* 18: 1001-16.
- Patterson, T. C., and M. E. Moseley. 1968. Pre-ceramic and early ceramic cultures of the central coast of Peru. *Nawpa Pacha* 6: 115-33.
- Perl, T. M., L. Bedard, T. Kosatsky, et al. 1990. An outbreak of toxic encephalopathy caused by eating mussels contaminated with domoic acid. *New England Journal of Medicine* 322: 1775-80.
- Perl, T. M., R. Remis, T. Kosatsky, et al. 1987. Intoxication following mussel ingestion in Montreal. *Canada Diseases Weekly Report* 13: 224-6.
- Petroff, I. 1884. *Alaska: Its population, industries, and resources*. Washington, D.C.
- Polo, J. A. 1977. *Nombres vulgares y usos de las algas en el Perú*, Serie de divulgación, Universidad Nacional Mayor de San Marcos, Museo de Historia Natural Javier Prado, Departamento de Botánico, No. 7. Lima.
- Porsild, A. E. 1953. Edible plants of the Arctic. *Arctic* 6: 15-34.
- Pozorski, S. G. 1979. Prehistoric diet and subsistence of the Moche Valley, Peru. *World Archaeology* 11: 163-84.
- Ragan, M. A. 1981. Chemical constituents of seaweeds. In *The biology of seaweeds*, ed. C. S. Lobban and M. J. Wynne, 589-626. Berkeley, Calif.
- Raymond, J. S. 1981. The maritime foundation of Andean civilization: A reconsideration of the evidence. *American Antiquity* 46: 806-21.
- Reagan, A. B. 1934. Plants used by the Hoh and Quilente Indians. *Transactions of the Kansas Academy of Science* 37: 55-71.
- Robbs, P. G., J. A. Rosenberg, and F. A. Costa. 1983. Contenido vitamínico de *Scenedesmus quadricauda*. II. Vitamin B-12. *Revista Latinoamericana de Microbiología* 25: 275-80.
- Ryther, J. H., J. A. De Boer, and B. E. Lapointe. 1978. Cultivation of seaweed for hydrocolloids, waste treatment and biomass for energy conversion. *Proceedings of the Ninth International Seaweed Symposium*, ed. A. Jensen and R. Stein, 1-16.
- Salcedo-Olavarrieta, N., M. M. Ortega, M. E. Marin-Garcia, and C. Zavala-Moreno. 1978. Estudio de las algas comestibles del Valle de México. III. Análisis comparativo de aminoácidos. *Revista Latinoamericana de Microbiología* 20: 215-17.
- Savageau, C. 1920. *Utilisation des algues marines*. Paris.
- Schönfeld-Leber, B. 1979. Marine algae as human food in Hawaii, with notes on other Polynesian islands. *Ecology of Food and Nutrition* 8: 47-59.
- Scott, R. 1954. Observations on the iodo-amino acids of marine algae using iodine-131. *Nature* 173: 1098-9.
- Shaw, R. 1966. The polyunsaturated fatty acids of microorganisms. *Advances in Lipid Research* 4: 107-74.
- Shiraishi, Y., T. Shirotori, and E. Takahata. 1973. Determination of polycyclic aromatic hydrocarbon in foods. II. 3,4-Benzopyrene in Japanese foods. *Journal of the Food Hygiene Society of Japan* 14: 173-8.
- Shirotori, T. 1972. Contents of 3,4-benzopyrene in Japanese foods. *Tokyo Kasei Daigaku Kenkyu Kiyo* No. 12: 47-53.
- Simoons, F. J. 1991. *Food in China*. Boca Raton, Fla.
- Skvortzov, V. B. 1919-22. The use of Nostoc as food in N. China. *Royal Asiatic Society of Great Britain and Ireland* 13: 67.
- Smith, D. G., and E. G. Young. 1955. The combined amino acids in several species of marine algae. *Journal of Biochemistry* 217: 845-53.
- Smith, H. M. 1933. An edible mountain-stream alga. *Siam Society of Bangkok. Natural History Supplement* 9: 143.
- Spalding, B. J. 1985. The hunt for new polymer properties. *Chemical Weekly* 136: 31-4.
- Subba Rao, G. N. 1965. Uses of seaweed directly as human food. *Indo-Pacific Fisheries Council Regional Studies* 2: 1-32.
- Subbulakshmi, G., W. E. Becker, and L. V. Venkataraman. 1976. Effect of processing on the nutrient content of the green alga *Scenedesmus acutus*. *Nutrition Reports International* 14: 581-91.
- Tiffany L. H. 1958. *Algae, the grass of many waters*. Springfield, Ill.
- Tilden, J. E. 1929. The marine and fresh water algae of China. *Lingnan Science Journal* 7: 349-98.
- Tipnis, H. P., and R. Pratt. 1960. Protein and lipid content of *Chlorella vulgaris* in relation to light. *Nature* 188: 1031-2.
- Trubachev, N. I., I. I. Gitel'zon, G. S. Kalacheva, et al. 1976. Biochemical composition of several blue-green algae and *Chlorella*. *Prikladnaya Biokhimiya Microbiologia* 12: 196-202.
- Tseng, C.-K. 1933. *Gloiopeltis* and other economic seaweeds of Amoy, China. *Lingnan Science Journal* 12: 43-63.
1935. Economic seaweeds of Kwangtung Province, S. China. *Lingnan Science Journal* 14: 93-104.
1983. *Common seaweeds of China*. Beijing.
1987. Some remarks on kelp cultivation industry of China. In *Seaweed cultivation for renewable resources*, ed. K. T. Bird and P. H. Benson, 147-53. Amsterdam.
1990. The theory and practice of phycoculture in China. In *Perspectives in phycology*, ed. V. N. Rajarao, 227-46. New Delhi.
- Turner, N. J. 1974. Plant taxonomic systems and ethnobotany of three contemporary Indian groups of the Pacific Northwest (Haida, Bella Coola, and Lillooet). *Syesis* 7: 1-104.
1975. *Food plants of British Columbia Indians. Part I - Coastal peoples*, Handbook No. 34. Victoria.
- Turner, N. J., and M. A. M. Bell. 1973. The ethnobotany of the southern Kwakiutl Indians of British Columbia. *Economic Botany* 27: 257-310.
- Tutour, B. le. 1990. Antioxidation activities of algal extracts, synergistic effect with vitamin E. *Phytochemistry* 29: 3757-65.
- Ueyanagi, J., R. Nawa, Y. Nakamori, et al. 1957. Studies on the active components of *Digenea simplex* Ag. and related compounds. XLVIII. Synthesis of alpha-kainic acid. *Yakugaku Zasshi* 77: 613.
- Velasquez, G. T. 1972. Studies and utilization of the Philippine marine algae. In *Proceedings of the Seventh International Seaweed Symposium*, ed. K. Nisizawa, 62-5. New York.
- Venkataraman, L. V., W. E. Becker, and T. R. Shamala. 1977. Studies on the cultivation and utilization of the alga *Scenedesmus acutus* as a single cell protein. *Life Sciences* 20: 223-34.
- Waaland, J. R. 1981. Commercial utilization. In *The biology of*

- seaweeds, ed. C. S. Lobban and M. J. Wynne, 726-41. Berkeley, Calif.
- Watanabe, A. 1970. Studies on the application of Cyanophyta in Japan. *Schweizerische Zeitschrift für Hydrologie* 32: 566-9.
- Wester, P. J. 1925. *The food plants of the Philippines*, Bulletin No. 39. Manila.
- Wood, B. J. B. 1974. Fatty acid and saponifiable lipids. In *Algal physiology and biochemistry*, ed. W. D. P. Stewart, 236-65. Berkeley, Calif.
- Wood, E. J. F. 1965. *Marine microbial ecology*. London.
- Xia, B., and I. A. Abbott. 1987. Edible seaweeds of China and their place in the Chinese diet. *Economic Botany* 41: 341-53.
- Yacovlev, E., and F. L. Herrera. 1934-5. El mundo vegetal de los antiguos peruanos. *Revista Museo Nacional* 3: 241-322, 4: 29-102.
- Yacovlev, E., and J. C. Muelle. 1934. Un fardo funerario de Paracas. *Revista Museo Nacional* 3: 63-153.
- Yamamoto, T., T. Yamaoka, S. Tuno, et al. 1979. Microconstituents in seaweeds. *Proceedings of the Seaweed Symposium* 9: 445-50.
- Yanovsky, E. 1936. *Food plants of the North American Indians*. United States Department of Agriculture Miscellaneous Publication No. 237. Washington, D.C.
- Zaneveld, J. S. 1950. The economic marine algae of Malaysia and their applications. *Proceedings of the Indo-Pacific Fisheries Council*, 107-14.
1951. The economic marine algae of Malaysia and their applications. II. The Phaeophyta. *Proceedings of the Indo-Pacific Fisheries Council*, 129-33.
1955. *Economic marine algae of tropical South and East Asia and their utilization*. Indo-Pacific Special Publications, No. 3. Bangkok.
1959. The utilization of marine algae in tropical South and East Asia. *Economic Botany* 13: 90-131.
- Zimmermann, U. 1977. Cell turgor pressure regulation and turgor-mediated transport processes. In *Integration of activity in the higher plant*, ed. D. H. Jennings, 117-54. Cambridge and New York.
1978. Physics of turgor and osmoregulation. *Annual Review of Plant Physiology* 29: 121-48.
- Zimmermann, U., and E. Steudle. 1977. Action of indoleacetic acid on membrane structure and transport. In *Regulation of cell membrane activities in plants*, ed. C. Marre and O. Cifferi, 231-42. Amsterdam.
- Zimmermann, U., E. Steudle, and P. I. Lelkes. 1976. Turgor pressure regulation in *Valonia utricularis*: Effect of cell wall elasticity and auxin. *Plant Physiology* 58: 608-13.

In terms of their common botanical characteristics, alliums are mainly herbaceous plants, incorporating various underground storage structures made up of rhizomes, roots, and bulbs. The foliar leaves alternate, often sheathing at the base to give the superficial impression that they originate from an above-ground stem. As a rule, the flower cluster, or inflorescence, is umbrella-like, with all the flower stalks radiating from the same point (umbel); the flowers are pollinated by insects; the fruits take the form of a capsule or berry; and the seeds are numerous and endospermic.

This genus is placed in the lily family. Most, but not all, of the species possess the pungent odor typical of onion and garlic. In addition to alliums, species of *Ipheion*, *Adenocalymma*, *Androstephium*, *Esperocalis*, *Talbaghia*, *Nectarosiodum*, *Nilula*, and, possibly, *Descurainia* produce pungent odors (Fenwick and Hanley 1985a).



II.C.2 🌱 The *Allium* Species (Onions, Garlic, Leeks, Chives, and Shallots)

The genus *Allium* comprises more than 600 different species, which are found throughout North America, Europe, North Africa, and Asia. Approximately 30 species have been regularly used for edible purposes (although less than half of these are subject to cultivation), with the most important being onions, garlic, leeks, chives, and shallots.

Onions

History

Antiquity. The onion (*Allium cepa*) may have originated in Persia (Iran) and Beluchistan (eastern Iran and southwestern Pakistan). But it is also possible that onions were indigenous from Palestine to India. They have been known and cultivated for many thousands of years and no longer grow wild. Their range – virtually worldwide – now includes China, Japan, Europe, northern and southern Africa, and the Americas (Hedrick 1972).

The consumption of onions is depicted in the decoration of Egyptian tombs dating from the Early Dynasty Period, c. 2925–c. 2575 B.C. During the Old Kingdom, c. 2575–c. 2130 B.C., onions were used as religious offerings. They were put on altars and, as is known from mummified remains, were employed in preparing the dead for burial (placed about the thorax and eyes, flattened against the ears, and placed along the legs and feet and near the pelves). Flowering onions have often been found in mummies' chest cavities (Jones and Mann 1963). If Juvenal (Roman poet and satirist, c.A.D. 55–127) is to be believed, a particularly delicious onion was worshiped as a god by certain groups in ancient Egypt (Hyams 1971).

The Greek historian Herodotus reported that onions, along with radishes and garlic, were a part of the staple diet of the laborers who built the Great Pyramid at Giza (2700–2200 B.C.) (Jones and Mann 1963). Egyptian onions were said to be mild and of an excellent flavor, and people of all classes (save for priests, who were prohibited from eating them) consumed them both raw and cooked (Hedrick 1972).

In Sumeria (southern Iraq), onions were grown and widely used for cooking 4,000 years ago (Fenwick and Hanley 1985a), and both garlic and onions have been unearthed at the royal palace at Knossos in Crete (Warren 1970). Minoan voyages from the eastern Mediterranean (2000–1400 B.C.) doubtless helped in dispersing alliums from that region.

The ancient Greek physician Hippocrates (460–375 B.C.) wrote that onions were commonly eaten, and Theophrastus (c. 372–287 B.C.) listed a number of onion varieties, all named after places where they were grown: Sardian (from western Turkey), Cnidian (from southern Turkey), Samothracian (from a Greek island in the northeast Aegean), and Setanian (possibly from Sezze or Setia in central Italy) (Jones and Mann 1963; Warren 1970).

Asia. According to Charaka, a Hindu physician of the second century A.D., the onion (as in ancient Egypt) was thought not to be a suitable food for persons pursuing the spiritual life. Thus, the onion was taboo for orthodox Brahmins, Hindu widows, Buddhists, and Jains (Hyams 1971).

In China, the fifth-century treatise on agriculture, *Ch'i-min-yao-shu* (Essential Arts for the People) by

Chia Ssu-hsieh, described the cultivation of *ts'ung*, or spring onion (*Allium fistulosum* L.), along the Red River valley (Li 1969). Infusions of onion have long been used in China as a treatment for dysentery, headache, and fever (Hanley and Fenwick 1985).

In 1886, Kizo Tamari, a Japanese government official, stated that in his country, onions did not have globular bulbs but were grown like celery and had long, white, slender stalks (Hedrick 1972). Interestingly, some modern Japanese communities forbid the cultivation, but not the consumption, of the spring onion (Kuroda 1977).

Europe. Columella (Lucius Junius Moderatus Columella), a Spanish-born Roman agriculturalist of the first century A.D., wrote of the *Marsicam*, which the country people called *unionem* (a term that may be the origin of the English word "onion" and the French *oignon*) (Fenwick and Hanley 1985a). Columella's contemporary, the Roman gourmet Apicius (Marcus Gavius Apicius), created several recipes that employed onions, although he viewed the vegetable as a seasoning rather than a food in its own right (Fenwick and Hanley 1985a).

Writing at about the same time as Apicius, Dioscorides (a Greek military physician) described onions as long or round and yellow or white, and provided detailed discussions of the uses of garlic, onion, and other alliums as medicinal plants (Jones and Mann 1963; Warren 1970).

Still another contemporary, Pliny the Elder, told his readers that the round onion was the best and that red onions were more highly flavored than white. His *Natural History* described six types of onions known to the Greeks: Sardian, Samothracian, Alsidenian, Setanian, the split onion, and the Ascalon onion (shalot). Pliny claimed onions to be effective against 28 different diseases (Fenwick and Hanley 1985a). Then, later on in the first millennium, Palladius (Rutilius Taurus Aemilianus Palladius), a Roman agriculturist in about the fourth century (or later), gave minute directions for culturing onions and comprehensively described their cultivation (Hedrick 1972).

By the beginning of the second millennium, many accounts of foodstuffs were being penned by monks. For example, Peter Damian (1007–72), the founder of a reformed congregation of Benedictines in central Italy, indicated that he permitted a moderate dish of edible roots, vegetables – mostly onions, leeks, and chickpeas – and fruit on days when fasting was not prescribed. These meals were eaten both cooked and uncooked, and sometimes enlivened with oil on special feast days (Lohmer 1988).

The German Dominican monk and scientist Albertus Magnus (1193–1280) did not include onions in his lists of garden plants, but garlic and leeks were represented there, suggesting the esteem in which they were held. Onions, however, were exotic plants understood to have favorable effects on fertility by

generating sperm in men and lactation in women (Mauron 1986).

By the sixteenth century, onions were no longer exotic. The Portuguese physician Amatus Lusitanus (1511–68) wrote that they were the commonest of vegetables, occurring in red and white varieties, and had sweet, strong, and intermediate qualities. The German physician and poet Petrus Laurembergius (1585–1639) described some of these qualities, writing of the Spanish onion as oblong, white, large, and excelling all others in sweetness and size; he further reported that at Rome, the Caieta variety brought the highest price, but at Amsterdam the most valued variety was the St. Omer.

A nutritional revolution occurred in the nineteenth century, when food items previously monopolized by the upper classes became available to all. The defeat of scurvy began with the addition to the diet of potatoes and onions, which were progressively supplemented with other legumes and fruits. By the middle of the nineteenth century, deaths from tuberculosis were in decline. Among other things, this was the product of the continuing introduction into the diet of foods containing vitamins A, C, and E, as well as meat and fish, which provide the amino acids vital to the creation of antibodies (Knapp 1989).

The Americas. It is probable that the men of Christopher Columbus's crews sowed onions on Hispaniola as early as 1494, and Hernando Cortés reportedly encountered onions, leeks, and garlic on his march to Tenochtitlan in 1519. Interestingly, native Mexicans apparently had a lengthy acquaintance with this Eurasian plant, because it had a name – *xonacatl* (Hedrick 1972).

Onions were mentioned as cultivated in Massachusetts as early as 1629, in Virginia in 1648, and at Mobile, Alabama, in 1775. By 1806, six varieties of onions were listed as esculents in American gardens. In 1828, the potato onion (multiplier onion) was described as a vegetable of late introduction into the United States, and by 1863, 14 varieties were mentioned (Hedrick 1972; Toma and Curry 1980).

Recent production statistics. The major producers of dry onions (cured but not dehydrated) in 1996 were (in metric tons) China (9,629,895), India (4,300,000), the United States (2,783,650), Turkey (1,900,000), Japan (1,262,000), Iran (1,199,623), Pakistan (1,097,600), and Spain (1,018,100), and total world production was 37,456,390 metric tons. The major producers of spring onions (scallions) were Mexico (702,478), Korea (553,000), Japan (545,600), China (282,329), Turkey (230,000), and Nigeria (200,000), and world production was 3,540,595 metric tons.

The major exporters of onions in 1996 were the Netherlands, India, the United States, Argentina, Spain, Mexico, Turkey, and New Zealand. Major importers

were Germany, the Russian Federation, Brazil, Malaysia, Saudi Arabia, and the United Arab Emirates.

Horticulture and Botany

Botany. The common onion is known only in cultivation. Propagation is usually by division, although some strains may also produce seed. Spring onions, used mainly in salads, are always grown from seed and harvested young. Pickling onions are made small by planting them close together (Traub 1968).

Onion leaves are the thickened bases of the normal leaves from the previous season. The bulb is composed of fleshy, enlarged leaf bases; the outermost leaf bases do not swell but become thin, dry, and discolored, forming a covering (Fenwick and Hanley 1985a). The onion usually flowers in the spring. Honeybees prefer the nectar of *A. cepa* to that of *A. fistulosum* (green onions) (Kumar and Gupta 1993).

Cultivation. Two crops of onions are grown each year in the United States. That of the spring is grown in Arizona, California, and Texas. The summer crop, much larger, consists of nonstorage produce, mostly from New Mexico, Texas, and Washington, and storage produce, grown mainly in Colorado, Idaho, Michigan, New York, Oregon, and Washington (Fenwick and Hanley 1985a).

Onions grow best in fine, stone-free, well-irrigated soils. Their comparatively thick, shallow roots require high levels of nitrogen, phosphorous, and potassium for maximum yield. The onion does not compensate for water stress and is sensitive to salinity. Flavor maturation and bulb development are affected by high temperature, high light intensity, soil moisture, and nitrogen deficiency (Brewster 1977a, 1977b). Increased flavor strength is associated with higher levels of applied sulfate (Platenius 1941; Kumar and Sahay 1954). Bulb formation depends upon increased daylength, but the daylength period required varies greatly between cultivars (Austin 1972).

Intercropping and rotation. Onions are the highest-yielding and most profitable inter- or border-crop for finger millet (*Eleusine coracana*) wherever it is grown (Siddeswaran and Ramaswami 1987). With tomatoes, planting four rows of onions (15 centimeters apart) between two rows of tomatoes has provided a 36 percent higher tomato equivalent yield without significantly affecting the number, average weight, and marketable yield of the tomato fruits. The tomato and onion combination also provides the highest net returns and maximum profit (Singh 1991).

Harvesting. Mature onions suffer lower storage losses than those harvested early. As onions reach maturity, the tops soften just above the bulb junction and cause the leaves to fall over. They are usually harvested when most of the plants are in this state

(Kepka and Sypien 1971; Rickard and Wickens 1977). Harvesting methods depend on the size of the crop, the climate, and regional or national practices (Jones and Mann 1963).

After harvesting, unless the crop is immediately sent to market, curing is necessary. The purpose of curing is to dry the skins and the top of the onion, forming an effective barrier against attack by microorganisms and, at the same time, minimizing the weight loss of the bulb. The onion is cured when the neck is tight, the outer scales are dry, and 3 to 5 percent of the original bulb weight is lost (Thompson, Booth, and Proctor 1972).

Curing can be natural or artificial. Windrowing, the traditional method in Britain, leaves the onions in the field, with the leaves of one row protecting the bulbs in the next. Direct exposure of the bulbs to the sun, especially under moist conditions, may lead to fungal damage (Thamizharasi and Narasimham 1993). In many countries, the onions are braided into bunches and hung up to dry (Thompson 1982).

Artificial curing techniques include forced heated air, vacuum cooling, cold storage, and infrared irradiation (Buffington et al. 1981). A small-capacity dryer has been developed in India (Singh 1994).

Storage. In addition to effective harvesting and curing, the critical factors for successful storage are cultivar type, storage conditions, and storage design. Losses from rotting and sprouting are more important than those from desiccation. Onions best suited for long-term storage (up to six months) usually have high amounts of dry matter and soluble solids, a long photoperiod during bulb maturation, and strong pungency. Red onions store better than white ones (Jones and Mann 1963; Thompson et al. 1972).

Temperature and relative humidity are the most important factors in storage conditions. Cold storage produces the best results but is not feasible in the tropics, where high-temperature storage may be effective, because dormancy is longer at 0° C and at 30° C than in between (10–15° C). Humidity should be about 70 to 75 percent (Robinson, Browne, and Burton 1975). Controlled-atmosphere storage losses depends on the quality and condition of the crop prior to storage (Adamicki and Kepka 1974).

In storage design, aeration is important for curing the onions and ventilating the heap. Consequently, slatted floors (or a similar layout, so that air can move through the bulbs from below) are employed. The onions are positioned so that air flows throughout the heap; otherwise, the moist, warm air retained in the middle leads to sprouting or rotting. The heaps should not be more than 8 feet high, and – especially where temperature, aeration, and humidity control are difficult – shallow heaps are recommended (Hall 1980).

Gamma irradiation is an effective inhibitor of sprouting in onion and garlic bulbs. Studies have

shown that eating irradiated onions does not harmfully affect animals or their offspring, but irradiation can cause discoloration, may not affect rotting, and may make onions more susceptible to aflatoxin production (Van Petten, Hilliard, and Oliver 1966; Van Petten, Oliver, and Hilliard 1966; Priyadarshini and Tulpule 1976; Curzio and Croci 1983).

Pathogens and Pests

Fungi

DOWNY MILDEW. Downy mildew (*Peronospora destructor* [Berk.] Casp.) was first reported in England in 1841. It is now widespread and particularly prevalent in cool, moist climates such as the coastal regions bordering the North Sea in Britain and those of the northwestern (Washington, Oregon, and California) and northeastern (New York and New England) United States. This fungus attacks onions, garlic, leeks, and chives alike. Early infection may kill young plants, and survivors can be dwarfed, pale, and distorted. Later infection causes chlorosis and yellowing of the leaves and stems. Some plants may be systemically infected and, if used for propagation, can serve as sources of inoculum in the seed crop.

When infected, the bulb tissue tends to soften and shrivel, and the outer fleshy scales become amber-colored, watery, and wrinkled. Underlying scales may appear healthy and yet be heavily infected. The fungus commonly overwinters in young autumn-sown onions whose leaves have been infected by neighboring summer crops.

Downy mildew can be controlled by growing onions on uncontaminated land without adjacent diseased crops. Good-quality, noninfected onions should be used, and planting should be done on open, well-drained land (Fenwick and Hanley 1985a).

WHITE ROT. White rot (*Sclerotium cepivorum* Berk.) was first noted in mid-nineteenth-century England and, like downy mildew, infects all the alliums under scrutiny in this chapter. The fungal attack is favored by dry soil and cool conditions. It develops rapidly between 10° C and 20° C and is inhibited above 24° C, although periods of dry weather can lead to devastating attacks in the field. When young plants are attacked, the disease spreads rapidly. External signs are yellowing and necrosis of leaf tips. Roots and bulbs are also affected. The bulb scales become spongy, are covered with fluffy white mycelium, and develop black sclerotia. The fungus appears to overwinter as sclerotia, and, in fact, the sclerotia may survive 8 to 10 years or more in the absence of host plants. Growing seedlings and sets for transplanting in noninfected soil, and the use of long rotations, are of some benefit in controlling this fungus. Chemical treatment with mercuric chloride, lime, and 2,6-dichloro-4-nitroaniline have also proven effective (Fenwick and Hanley 1985a).

ONION SMUDGE. Common now in Europe and the United States, onion smudge (*Collectotrichum circinans* [Berk.] Vogl.) was first reported in England in 1851. It affects mainly white varieties of onion but has been reported in shallots and leeks. It is confined to the necks and scales, where it causes blemishes, reducing the market value of the crop. Rarely attacking the active growing parts of the plant, it is confined on colored onions to unpigmented areas on the outer scales of the neck. Onion smudge requires warm, moist conditions (10° C to 32° C, optimum 26° C). Conidia, or fungal spores, are produced abundantly and are scattered by spattering rain. With suitable conditions, a conidial spore will germinate within a few hours. Pungent onions resist smudge better than mild ones. Crop rotation, good husbandry, and carbamate sprays can minimize the damage. Drying the onions in hot air may be necessary, and curing under dry, well-ventilated conditions is important (Fenwick and Hanley 1985a).

ONION SMUT. Probably originating in the United States in the late nineteenth century and first reported in Britain in 1918, onion smut (*Urocystis cepulae* Frost) attacks bulb and salad onions as well as leeks, shallots, chives, and garlic. Infection occurs from two to three weeks after sowing, and a high percentage of the infected plants subsequently die. Elongated, leaden streaks discolor the scales and the growing leaves, which can also become thickened and malformed. The streaks develop into smut sori, which rupture and release spores that can survive up to 20 years in soil. Measures to control this fungus include avoiding infected areas, pelleting seed with hexachlorobenzene, dusting with thiram or ferbane, and applying fungicides (Fenwick and Hanley 1985a).

NECK ROT. Caused by three different species of *Botrytis*, neck rot is probably the most widely distributed and most destructive disease of onions in storage. It was first reported in Germany (1876) and then in the United States (1890) and Britain (1894). Infection occurs in the field but is usually not noticed until harvesting occurs. The first signs are a softening of bulb scales and the development of sunken brown lesions; a definite border between fresh and diseased tissue can be seen. The bulb desiccates and collapses. If the onions are stored in moist conditions, a secondary spread may take place. Infection occurs primarily from spores dispersed by wind or water before, during, or after harvest. White onions seem more susceptible than yellow or colored varieties, and pungent onions are less affected than mild-flavored varieties. Practical controls are thin sowing, careful handling during harvest, and providing optimal storage conditions. Zineb and other chemicals, including carbamate sprays, reduce infection, and in recent years, benomyl seed dressings have also been used effectively (Fenwick and Hanley 1985a).

Bacteria

SOFT ROT. The soft rot pathogen (*Erwinia carotovora*) enters onions through wounds that occur during harvest, transportation, and storage, or in the necks of uncured or slow-curing varieties. The infection usually starts at the bulb neck, with external signs of sponginess and a foul-smelling exudate from the neck when the bulb is squeezed. Soft rot occurs most commonly in humid weather and is transported by the onion maggot, which is itself contaminated by the rotting vegetation it consumes and, consequently, lays eggs carrying the bacteria. Control involves avoiding damage to the bulbs during and after harvest, drying them thoroughly and rapidly, using the lowest practicable storing temperature, and eliminating all damaged bulbs (Fenwick and Hanley 1985a; Wright, Hale, and Fullerton 1993). Also important is moving bulbs under cover and drying them if wet weather is expected during field-curing (Wright 1993).

Viruses. "Aster yellows," spread by the six-spotted leafhopper (*Macrostelus facifrons*), is an important viral disease of onion as well as of carrot, barley, celery, endive, lettuce, parsley, potato, and salsify. Yellowing young leaves are followed by the appearance of yellowed shoots; and roots become small and twisted. Control measures consist of reducing or eradicating the leafhopper population where aster yellows is prevalent (Fenwick and Hanley 1985a).

Nematodes. The bulb and stem nematode (*Ditylenchus dipsaci* [Kuhn] Filipjer) is widespread in the Mediterranean region but has also been found on onions and garlic in the United States, on onions in Brazil and England, and on onions and chives in Holland. It causes a condition known as "onion bloat." Dead plant tissue can contain dormant nemas, which are probably an important source of infestation. Chloropicren/steam fumigation and other treatments have proven effective, but bromine-containing nematocides should be avoided. Both onion and garlic are bromine-sensitive and will not produce good crops for up to 12 months if bromine residues are present in the soil.

Ditylenchus dipsaci is widespread in southern Italy, where it reproduces on several wild and cultivated plant species. Among vegetables, the most severely damaged are onion and garlic, but broad bean, pea, and celery also suffer damage. In the Mediterranean area, the nematode mainly infects host plants from September to May, but reproduction is greatest in October, November, March, and April, when soil moisture, relative humidity, and temperatures are optimal. Symptoms of nematode attack are apparent in the field from late February to April and in nurseries during October and November. As a result, early crops are damaged more than late crops. Nematodes survive in the soil and in plant residues. However, seeds from infested plants, except those of

broad bean and pea, have rarely been found to harbor nematodes. The use of seeds, bulbs, and seedlings free of nematodes is a prerequisite for successful crop production. Cropping systems, soil treatments with fumigant and nonvolatile nematocides, and soil solarization of infested fields are recommended for effective and economic nematode control (Greco 1993).

Insects. Although many insects can attack onions, the two major culprits are the onion thrip (*Thrips tabaci* Lind.) and the onion maggot, the larval stage in the development of the onion fly (*Hylemya antiqua* Meig.). The onion thrip punctures leaves and sucks the exuding sap, leaving whitish areas on the leaves. Infestation is worse in very dry seasons and can often lead to the destruction of entire crops. Effective chemicals are available to control this pest, and results have shown that a 40 percent bulb-yield reduction occurs on nontreated plots as compared with treated ones (Domiciano, Ota, and Tedardi 1993).

The onion maggot is a pest of considerable economic importance. Both the fly and its eggs are carriers of the soft rot pathogen *E. carotovora* Holland. The adult female lays 30 to 40 eggs in the soil around the onion plant or on the onion itself, especially where plants are damaged, decaying, or already infected with larvae. Good husbandry, the destruction of onion waste, and chemicals such as aphidan, EPBP, fensulfothion, fonofos, malathion, or phoxim are used to control the onion fly and its offspring (Fenwick and Hanley 1985a).

Processing

Dehydrated onion pieces. After grading and curing, onions are peeled using lye or the flame method, whereby the roots and outer shell are burnt off in an oven, and the charred remnants are removed by washing. Next, the onions are sliced by revolving knives and dried by hot air forced upward through holes in the conveyor belt. For good storage and acceptable flavor stability, residual moisture content is about 4 to 5 percent. Moisture content can be reduced to the desired level in one to two hours (Gummery 1977). The onion pieces may then be used as such or converted into granules and flakes (or powder). Dehydrated onion pieces are widely employed in the formulation of sausage and meat products, soups, and sauces (Hanson 1975; Pruthi 1980).

Onion powder. Onion powder is used in cases where onion flavor is required but the appearance and texture of onions are not, as in dehydrated soups, relishes, and sauces. Onion powder is made by grinding dehydrated onion pieces or by spray-drying. For spray-drying, onions are washed free of debris, rinsed, and blended to a puree. Dextrose (30 to 40 percent by weight) is added, and the mixture spray-dried at temperatures below 68° C. It can be dried in four minutes at 65° C to 68° C. The treatment destroys all patho-

genic bacteria while reducing the bacterial population, and the end product has excellent keeping properties (Gummery 1977).

Onion oil. Distillation of minced onions that have stood for some hours produces onion-essential oil. The oil is a brownish-amber liquid that contains a complex mixture of sulfur and other volatiles. The oil has 800 to 1,000 times the odor of a fresh onion, and its price may be 1,000 times more expensive as well. It is used for its solubility, lack of color, and strong aroma. However, onion oil cannot be standardized because its composition depends on the onion variety, ecological conditions, season, and processing (Heath 1981).

Onion juice. Onion juice is produced by expressing the bulbs, flash-heating the liquor obtained to a temperature of 140° C to 160° C, and immediately cooling it to 40° C. Next, the juice is carefully evaporated to approximately 72 to 75 percent dry matter to preserve it without chemical additives. The concentrated juice is pale brown in color and possesses a strong, fresh onion odor. Further evaporation to 82 to 85 percent solids darkens the product and gives it a cooked, toasted effect preferred by many. The sensory qualities are sometimes enhanced by returning the aromatic volatile condensate to the juice. The extract is often mixed with propylene glycol, lecithin, and glucose to yield an onion oleoresin that has a flavor 10 times that of onion powder and 100 times that of the original bulb (Heath 1981).

Onion salt. In the United States, onion salt is a mixture of dehydrated onion powder (18 to 20 percent), calcium stearate (an anticaking agent - 1 to 2 percent), and sodium chloride.

Pickled onions. Onions are pickled in a 10 percent salt solution and preserved in vinegar. Generally, silverskin or button onions are used because they give a translucent product with the desired firmness of texture. Lactic acid bacteria are the important fermentation organisms, and care must be taken to keep the solution at 10 percent salinity. Finally, the salt is leached from the onions with warm water, and the bulbs are placed in cold, spiced vinegar and stored in sealed glass jars (Fenwick and Hanley 1985a).

Nutrition

The nutritional content of onions varies by variety, ecological conditions, and climate. According to the Nutrition Data System of the University of Minnesota, 100 grams (g) (3.53 ounces or 0.44 cup) of onion provides 38 kilocalories of energy, 1.16 g of protein, 0.16 g fat, and 8.63 g of carbohydrate.

Using the standard of the *Recommended Dietary Allowances* (tenth edition) for a male between 18 and 25 years of age, approximately 100 g or one-half cup of

fresh onion provides 10.7 percent of the Recommended Dietary Allowance (RDA) of vitamin C and 9.5 percent of folacin. Onions are high in potassium (157 milligrams [mg]) and low in sodium (3 mg). They contain small amounts of calcium, copper, iron, magnesium, manganese, molybdenum, phosphorus, selenium, and zinc (Raj, Agrawal, and Patel 1980). Other trace elements in onion are germanium, chromium, and lithium. Onions have no vitamin A and only small amounts of alpha-tocopherol, delta-tocopherol, thiamine, riboflavin, niacin, pantothenic acid, and vitamin B₆.

In addition, 100 g of onions contain only small amounts of three fatty acids: saturated palmitic acid (0.02 g), monounsaturated oleic acid (0.02 g), and polyunsaturated essential linoleic acid (0.06 g). They have 2.1 g of dietary fiber, no starch, and 89.68 g of water. Sucrose (1.3 g), glucose (2.4 g), and fructose (0.9 g) are present. All essential amino acids are present in onions. Arginine (0.16 g), which increases during maturation (Nilsson 1980), and glutamic acid (0.19 g) are the most abundant.

Chemistry

The color of red onions is due to cyanidin glycosides, anthocyanins that contain glucose molecules (Fuleki 1971). With yellow onions, quercetin, a flavonoid, and its glycosides are responsible for the color of the dry scales. The outer scales of onions have been used in Germany for dyeing Easter eggs and household fabrics (Perkin and Hummel 1896; Herrmann 1958). The flavonoid content is usually greatest in the outer leaves and may act as a protection against predators (Tissut 1974; Starke and Herrmann 1976a).

The phenolic compounds catechol and protocatechuic acid are found in greater quantities in colored onions than in white onions. The presence of these compounds in the outer dried scales is a contributing factor to the greater resistance of these types to smudge and neck rot diseases and to fungi-causing wild and soft rots (Walker and Stahman 1955; Farkas and Kiraly 1962).

The most important nonstructural polysaccharide in onion is a group of fructose polymers called fructans. Fructose commonly forms chains of 3 to 10 molecules, with chains of 3 and 4 molecules being the most common. It is thought that these polymers are used for storage carbohydrates and osmoregulation during bulb growth and expansion (Darbyshire and Henry 1978; Goodenough and Atkin 1981).

Onions contain pectins with high methoxyl content and of the rapid-setting kind. Pectin is used in the preparation of jellies and similar food products and is used by veterinarians as an antidiarrheal (Alexander and Sulebele 1973; Khodzhaeva and Kondratenko 1983). Onions also contain several sterols. Beta-sitosterol, cycloartenol, and lophenol are the most common, followed by campesterol. Beta-sitosterol is used as an antihyperlipoproteinemic (Oka, Kiriya, and Yoshida 1974; Itoh et al. 1977).

Like garlic, onion has exhibited antioxidative activity, which can be increased by microwave heating or boiling. It has been shown that S-alkenyl cysteine sulfoxides are the most active components. Quercetin and other flavone aglycones also contribute to the total antioxidative capacities of onion and garlic extracts (Pratt and Watts 1964; Naito, Yamaguchi, and Yokoo 1981a, 1981b).

Onions produce thiamine propyldisulfide, which corresponds to the allithiamine formed in garlic from thiamine and alliin. Both compounds have been found effective against cyanide poisoning (Carson 1987).

Medicinal Use

Atherosclerotic. Onion is known to have a hypocholesterolemic effect, although not as strong as that of garlic (Bhushan et al. 1976). A study in China compared an onion-growing region to one without local onions. Both regions were similar in living standards, economic level, and dietary habits and customs. But people in the onion-growing region had a death rate from cardiovascular disease of 57 per 100,000 people, as compared with a cardiovascular-disease death rate in the other region of 167 per 100,000. The onion-growing region also had a significantly lower incidence of hypertension, retinal arteriosclerosis, hyperlipemia, and coronary artery disease (Sun et al. 1993).

Hypo- and hyperglycemic effects. A study has revealed that although a water extract of fresh or boiled onion did not affect fasting blood sugar in normal subjects, it did reduce the sugar levels in glucose-tolerance tests in a dose-dependent manner. From this result, it was suggested that onion has an antihyperglycemic effect instead of a hypoglycemic effect (Sharma et al. 1977).

The antihyperglycemic principle in onion has been tentatively identified as 2-propenyl propyl disulfide – a compound that has been found to lower the blood sugar and increase insulin levels but has not been observed to have any effect on free fatty-acid concentrations (Augusti 1974; Augusti and Benaim 1975). Another antihyperglycemic compound causing this effect is diphenylamine, found in onion and tea (Karawya et al. 1984).

Ill-effects of consumption. One problem with the consumption of onions is heartburn, but only among those predisposed to heartburn symptoms (Allen et al. 1990). Onions may also cause discomfort in people with ileostomies and children with Down's syndrome (Bingham, Cummings, and McNeil 1982; Urquhart and Webb 1985).

As early as 1909, cattle deaths were attributed to eating sprouting or decaying onions (Goldsmith 1909; Fenwick and Hanley 1985c). Clinical signs of the condition may include onion odor in breath and urine,

tainting of milk, diarrhea, staggering, and collapse. Provided that the illness has not reached an irreversible point, the symptoms (which develop with a week of onion feeding) may decline when the offending ingredient is removed from the diet. Treatment may also include injection of B-complex vitamins with penicillin-streptomycin (Gruhzit 1931; Farkas and Farkas 1974; Kirk and Bulgin 1979).

Garlic

History

Antiquity. Cultivated in the Middle and Far East for at least 5,000 years, garlic (*Allium sativum*) is believed to have originated from a wild ancestor in central Asia and is, possibly, native to western Tartary (Turkestan). At a very early period, garlic was carried throughout the whole of Asia (except Japan), North Africa, and Europe. In ancient China, Egypt, and India, garlic – like onions – was a highly prized foodstuff (Hedrick 1972; Hanley and Fenwick 1985).

In Egypt, the consumption of garlic is shown in tomb art dating from the Early Dynastic Period (c. 2925–2575 B.C.). The Codex Elbers, an Egyptian medical papyrus dating from around 1500 B.C., described 22 garlic preparations employed against a variety of complaints, including headache, bodily weakness, and throat disorders (Fenwick and Hanley 1985a).

The Bible (Num. 11:5) reports that after their Exodus from Egypt (about 1450 B.C.), the Israelites complained to Moses about the lack of garlic, among other things: “We remember the fish which we used to eat free in Egypt, the cucumbers and the melons and the leeks and the onions and the garlic.”

The Greeks, along with the Egyptians, regarded garlic as a defense against old age and illness, and athletes participating in the Olympic Games, (which began about 776 B.C.), regularly chewed it to improve stamina (Hanley and Fenwick 1985). Homer, the Greek poet from the eighth century B.C., worked garlic into his tales (Hedrick 1972), including a description of how Odysseus fended off Circe’s magic using as antidote a plant “having black root and milk white flower” (Fenwick and Hanley 1985a: 202). Tradition has it that this plant was wild garlic (Fenwick and Hanley 1985a).

Hippocrates (c. 460–370 B.C.) recommended garlic for pneumonia and suppurating wounds, but warned that it “caused flatulence, a feeling of warmth on the chest and a heavy sensation in the head; it excites anxiety and increases any pain which may be present. Nevertheless, it has the good quality that it increases the secretion of urine” (Jones and Mann 1963; Warren 1970; Fenwick and Hanley 1985a: 202).

Asia. Garlic was introduced into China between 140 and 86 B.C. The Chinese word for garlic, *suan*, is written as a single character, which often indicates the

antiquity of a word (Hyams 1971). A fifth-century Chinese treatise on agriculture (*Ch’i-min-yao-shu*) described the cultivation of *suan* along the Red River valley. Chinese leeks, shallots, and spring onions were also discussed, but garlic seems to have been the most important. In addition, *tse suan* – water garlic (*Allium nipponicum* L.) – was mentioned as both a pervasive weed and a cultivated plant (Li 1969). According to Marco Polo (c. A.D. 1254–1324), garlic was used as a complement to raw liver among the Chinese poor (Lucas 1966), and much mention is made of garlic in treatises written in China from the fifteenth to the eighteenth centuries (Hedrick 1972).

In India, an important fifth-century Sanskrit medical manuscript, the *Charaka-Sambhita*, based on sources from perhaps five centuries earlier, attributed widespread curative properties to both garlic and onion. It was claimed that they possessed diuretic properties, were beneficial to the digestive tract, were good for the eyes, acted as heart stimulants, and had antirheumatic qualities (Fenwick and Hanley 1985a).

In the Ayurvedic (Sanskrit) and Unani Tibb (Greco-Arabic) systems, garlic has been employed both as a prophylactic and as a cure for a variety of diseases, including arteriosclerosis, cholera, colic, dysentery, dyspepsia, gastric and intestinal catarrh, and typhoid. Duodenal ulcers, laryngeal tuberculosis, and lupus have all been treated with garlic juice, and garlic preparations have been given for bronchiectasis, gangrene of the lung, pulmonary phthisis, and whooping cough (Fenwick and Hanley 1985a).

Today the use of garlic is especially prevalent in Asia, where garlic-based antibiotics are used extensively to replace or complement more sophisticated drugs (Hanley and Fenwick 1985). In addition, in rural villages of Karnataka, in southwestern India, garlic is prescribed for lactating women (Rao 1985).

Europe. Garlic was regularly mentioned in European literature as well, especially for its medicinal benefits. The Roman poet Virgil (79–19 B.C.), for example, in his *Second Idyll* described how Thestylis used the juices of wild thyme and garlic as a prophylactic against snake bites (Warren 1970). A bit later, Pliny the Elder, in his *Natural History*, recommended that garlic be “placed when the moon is below the horizon and gathered when it is in conjunction” (Fenwick and Hanley 1985a: 200) to remove the plant’s pungent smell. He devised 61 garlic-based remedies for such conditions as hemorrhoids, loss of appetite, rheumatism, and ulcers (Jones and Mann 1963; Fenwick and Hanley 1985a).

The Romans apparently disliked garlic in general because of its strong scent, but it was fed to laborers to strengthen them and to soldiers to excite courage. The Romans also used garlic as a remedy for diabetes mellitus, and it is probable that it was similarly employed by the Egyptians and Greeks (Hanley and

Fenwick 1985). Carbonized garlic has been found at Pompeii and Herculaneum, which were destroyed in A.D. 79 (Meyer 1980).

The Greek military physician Dioscorides (A.D. 40–90) was clearly impressed with garlic, onion, and other alliums as medicinal plants. He advised garlic for baldness, birthmarks, dog and snake bites, eczema, leprosy, lice, nits, toothache, ulcers, and worms. He also suggested it as a vermifuge and diuretic and as a treatment for rashes and other skin disorders (Warren 1970; Fenwick and Hanley 1985a).

The cultivation of alliums in Western Europe is usually thought to have been stimulated by the Crusaders' contacts with the East in the eleventh, twelfth, and thirteenth centuries. However, much earlier, Charlemagne (742–814) had listed garlic in his *Capitulaire de Villis* and mentioned it as of Italian origin (Fenwick and Hanley 1985a). During medieval times, garlic was less appreciated for its taste than for its allegedly favorable effect on sexual potency and performance (Mauron 1986).

Presumably, however, the latter was of little interest to St. Hildegard (1098–1179), a German abbess, mystic, and scientific observer who continued the focus on garlic as medicine by specifically mentioning it in her *Physica* as a remedy against jaundice. The herbal doctors Paracelsus (Philippus Aureolus Paracelsus, 1493–1541) and Lonicerus (Adam Lonitzer, 1528–86) emphasized the antitoxic properties of garlic and its effectiveness against internal worms. At about the same time, Italian physician and botanist Matthioli (Pietro Andrea Mattioli, 1500–77) was recommending garlic against stomach chills, colics, and flatulence.

The word “garlic” is derived from the old English “gar” (meaning spear) and, presumably, refers to the garlic clove. Geoffrey Chaucer (c. 1342–1400) wrote of “Wel loved garleek, onyons and leekes” (Fenwick and Hanley 1985a: 200), and garlic's pungency was described by William Shakespeare. In *A Midsummer Night's Dream* (Act IV, Scene 1), Bottom tells his fellow actors to eat neither garlic nor onion, “for we are to utter sweet breath,” and in *Measure for Measure* (Act III, Scene 2), Lucio criticizes the Duke, who “would mouth a beggar, though she smell brown bread and garlic.” A contemporary of Shakespeare described King Henry IV of France as “chewing garlic and having breath that would fell an ox at twenty paces” (Fenwick and Hanley 1985a: 201).

Garlic's medicinal (and supposedly aphrodisiacal) powers were known in England in the sixteenth and seventeenth centuries, and the diarist Samuel Pepys (1633–1703) discovered that the custom in the French navy – to keep the sailors warm and prevent scurvy – was to issue garlic and brandy rations; the British Admiralty followed suit (Fenwick and Hanley 1985a).

At the turn of the nineteenth century, garlic in the form of inhalants, compresses, and ointments was

used by the citizens of Dublin against tuberculosis, and the medicinal use of garlic is still common in Bulgaria, Japan, and Russia, among other places (Petkov 1986). In Russia, garlic-based antibiotics are widely employed, and on one occasion, 500 tonnes of garlic were imported to combat an outbreak of influenza (Fenwick and Hanley 1985a).

The Americas. Garlic was introduced to the Americas by the Spaniards. In Mexico, Cortés (1485–1547) apparently grew it, and by 1604, it was said in Peru that “the Indians esteem garlic above all the roots of Europe” (Hedrick 1972). By 1775, the Choctaw Indians of North America (Alabama, Louisiana, and Mississippi) were cultivating garlic in their gardens, and at the turn of the nineteenth century, American writers mentioned garlic as among their garden esculents (Hedrick 1972).

Garlic is widely used today in Latin America as a medicine as well as a food. In Guatemala, for example, it is prescribed for vaginitis by traditional healers, health promoters, and midwives (Giron et al. 1988) and is also employed against helminthic infection, both alone and in conjunction with commercial drugs (Booth, Johns, and Lopez-Palacios 1993). Argentine folk medicine prescribes garlic for antimicrobial use (Anesini and Perez 1993), and in the mountains of Chiapas in southeastern Mexico, Indian shepherders use garlic and other alliums for veterinary purposes (Perezgrovas Garza 1990).

Production. The major producers of garlic in 1996 were (in metric tons) China (8,574,078), Korea (455,955), India (411,900), the United States (277,820), Egypt (255,500), and Spain (212,400), and world production was 11,633,800 metric tons. Major exporters in 1996 were China, Hong Kong, Singapore, Argentina, Spain, Mexico, and France, and major importers were Malaysia, Brazil, Indonesia, Singapore, the United Arab Emirates, Japan, the United States, and France.

In the United States, garlic production is confined mostly to California. Most of this crop is grown around the town of Gilroy, which calls itself the “garlic capital of the world” (Fenwick and Hanley 1985a).

Horticulture and Botany

Botany. Garlic is known only in its cultivated form but may be related to the wild *Allium longicaule* of central Asia. Garlic bulbs develop entirely underground, and the plant is either nonflowering or flowers in the spring. Its leaves are flat and rather slender; the stem is smooth and solid. The bulbs are composed of several bulbils (cloves) encased in the white or pink skin of the parent bulb. Each clove is formed from two leaves, the outer cylindrical one being protective and the inner one a storage organ for the bud (Traub 1968).

Cultivation. Although it grows in a wide variety of soils, garlic flourishes best in rich, deep loams with plentiful moisture. Before planting, the bulbs should be dried, treated (e.g., with benomyl) to reduce rotting, and exposed to temperatures between 0° C and 10° C for four to eight weeks to ensure bulb-ing. Bulbs usually form and enlarge with long days and temperatures above 20° C. Plant spacing affects the size of the bulbs. Italian workers consider a spacing of 40 to 50 per square meter desirable. Doubling this density increases the yield by 50 percent, but then the bulbs are smaller and more suitable for processing than for the fresh market (Tesi and Ricci 1982). When the tops become dry and bend to the ground, harvesting is generally done by hand, although it can be done mechanically. Curing is usually carried out in the ground or in well-ventilated structures, and the dried bulbs can be stored.

Proper curing enables garlic to store well without careful temperature control. The best results are achieved when the bulbs are dried 8 to 10 days at 20° C to 30° C, followed by a reduction of temperature to 0° C with air circulation. Under these conditions, garlic bulbs can be stored from 130 to 220 days, depending on variety and how they were grown (IOS 1983).

Also effective in garlic storage is the application of maleic hydrazide prior to harvest (Omar and Arafa 1979), and gamma irradiation prevents storage losses without an adverse effect on taste, flavor, pungency, or texture (Mathur 1963). For cold storage conditions, it is recommended that garlic be harvested, dried, and packed away from all other crops except onions (Tesi and Ricci 1982).

Pathogens and Pests

The common pests and pathogens of garlic are those discussed in the section about onions.

Processing

Dehydrated garlic. As already mentioned, most of the garlic produced in the United States (90 percent) is grown and processed near the town of Gilroy, California. Gilroy also has the largest dehydration plant in the world, and in this region, more than 60,000 tons annually are processed into 25 different kinds of flakes, salts, and granules.

Dehydrated garlic can contain five times the flavor of the fresh clove, and garlic powder is used extensively in the manufacture of spiced sausages and other foods. To maintain flavor character and prevent lumping and hardening, the powder must be stored free of moisture. Flavor deterioration of stored garlic powder is maximal at 37° C and minimal between 0° C and 2° C. At room temperature, the product is best stored in cans. The packaging of garlic powder (at 6 percent moisture content) in hermetically sealed cans is best of all (Singh, Pruthi, Sankaran, et al. 1959; Singh, Pruthi, Sreenivasamurthy, et al. 1959).

Garlic flavoring. The volatile oil content of garlic is between 0.1 and 0.25 percent. The reddish-brown oil from the distillation of freshly crushed garlic cloves is rich in 2-propenyl sulfides. Often the oil itself is too pungent for efficient manufacturing use, so garlic juice – obtained in a similar manner to onion juice – is employed. Concentrating the juice produces oleoresin garlic, a dark-brown extract with approximately 5 percent garlic oil. The oleoresin has uniformity, good handling, and good processing characteristics.

Nutrition

As with onions, the nutrient content of garlic changes with variety, ecological conditions, and climate. One hundred grams (3.53 ounces, or 0.44 of a cup) of garlic provides about 149 kilocalories of energy, 6.36 g of protein, 0.5 g of fat, and 33.07 g of carbohydrate (Nutrition Coordinating Center 1994).

In light of the RDA standard for males between 18 and 25 years of age, approximately one-half cup of fresh garlic (100 g) would provide them with 10.1 percent of the recommended dietary allowance of protein, 22.6 percent of calcium (181 mg), 17 percent of iron (1.7 mg), 19.1 percent of phosphorus (153 mg), 13.3 percent of copper (0.3 mg), 20.3 percent of selenium (14.2 mg), 52 percent of vitamin C (31.2 mg), 13.3 percent of thiamine (0.2 mg), 10.9 percent of pantothenic acid (0.6 mg), and 61.5 percent of vitamin B₆ (1.23 mg). Garlic is high in potassium (401 mg/100 g), low in sodium (17 mg/100 g), and contains small amounts of magnesium, manganese, molybdenum, and zinc (Pruthi 1980; Raj et al. 1980). Other trace elements in garlic are cobalt, chromium, lithium, nickel, titanium, and vanadium. Garlic contains no vitamin A or E but does have small amounts of riboflavin, niacin, and folacin (National Research Council 1989).

Garlic (100 g) contains only small amounts of four fatty acids: 0.09 g of saturated palmitic acid, 0.01 g of monounsaturated oleic acid, 0.23 g of polyunsaturated essential linoleic acid, and 0.02 g of polyunsaturated essential linolenic acid. It has 4.1 g of dietary fiber, 14.7 g of starch, and 58.58 g of water. Sucrose (0.6 g), glucose (0.4 g), and fructose (0.6 g) are present, as are the essential amino acids – arginine (0.63 g) and glutamic acid (0.8 g) are the most abundant, followed by aspartic acid (0.49 g) and leucine (0.31 g).

Chemistry

Nonflavor compounds. Garlic contains polymers of fructose with up to 51 fructose molecules (Darbyshire and Henry 1981). It also yields pectin. Garlic pectin content includes galactose, arabinose, galacturonic acid, and glucose. It has a much higher viscosity than onion pectin, as well as a lower setting temperature and a longer setting time (Alexander and Sulebele 1973; Khodzhaeva and Kondratenko 1983).

Sterols found in garlic are stigmasterol, B-sitosterol,

and campesterol (Oka et al. 1974; Stoianova-Ivanova, Tzutzułova, and Caputto 1980). Garlic also contains arachidonic and eicosapentaenic acids (Carson 1987).

Garlic has exhibited antioxidant activity in linoleic-acid and minced-pork model systems. This activity can be increased by microwave heating or boiling. It has been shown that S-alkenyl cysteine sulfoxides were the most active. Quercetin and other flavone aglycones also contribute to the total antioxidant capacities of onion and garlic extracts (Pratt and Watts 1964; Naito et al. 1981a, 1981b).

Allithiamin, discovered in the 1950s by Japanese researchers, is formed in garlic from thiamine and alliin and is absorbed faster in the intestinal tract than thiamine (Fujiwara 1976). Unlike thiamine, allithiamin is not degraded by thiaminase and appears more stable under conditions of heat (Hanley and Fenwick 1985). Allithiamin, which reacts with the amino acid cysteine to regenerate thiamine – yielding 2-propenylthiocysteine – has been found effective against cyanide poisoning (Carson 1987).

Flavor compounds. The first important studies on the composition of garlic oil were carried out by T. Wertheim in 1844 and 1845. While investigating the antibacterial properties of garlic in the 1940s, C. J. Cavallito and others discovered the thiolsulfinate alliin, the most important flavor component of fresh garlic (Carson 1987). This colorless oil is di(2-propenyl)thiolsulfinate. (In this chapter, 2-propenyl is used instead of allyl.) Alliin is probably the first thiolsulfinate isolated from natural sources (Carson 1987).

The compounds responsible for the flavor of alliums are produced from involatile precursors only when tissue maceration occurs. Gamma-glutamyl peptides, containing approximately 90 percent of garlic's soluble, organically bound sulfur, are present in significant amounts and may be the storage form of the flavor precursors (Virtanen 1965; Whitaker 1976). Under these circumstances, alkyl or alkenyl cysteine sulfoxides come into contact with an enzyme, alliinase, and hydrolysis occurs. The initially formed thiolsulfinate can break down to produce a range of organoleptically important sulfur compounds, including disulfides, trisulfides, higher sulfides, and thiols. The flavor properties of the different alliums depend on the types and amounts of these sulfur compounds (Hanley and Fenwick 1985).

Over 90 percent of the flavor-precursor content of garlic is located in the storage leaf (Freeman 1975). Alliin lyase is a major product of the storage bud (clove), accounting for 10 percent of its total protein. Deposits of alliinase are most pronounced around phloem tissue and are concentrated in the bundle sheaths. Little, if any, occurs in storage mesophyll that is not in contact with vascular bundles. This deposition in the clove may reflect the enzyme's role in protecting underground storage buds from decay and predation. Positioning near the phloem suggests that

alliin lyase, or compounds related to its activity, may be translocated to and from the clove during development (Ellmore and Feldberg 1994). Alliinase is present in most, if not all, members of the genus *Allium*, and is also found in *Albizia*, *Acacia*, *Parkia*, and *Lentinus* species.

Medicinal Use

Atherosclerotic. Medical claims for the efficacy of garlic against myriad complaints have been made for millennia and are still being made today as science continues to analyze the properties of this tasty vegetable and channel them to medical use.

The second-century Indian physician, Charaka, reported that onion and garlic prevented heart disease and acted as heart tonics (Fenwick and Hanley 1985c). Clots, which can cause strokes and heart attacks, are formed through the aggregation of platelets. Both garlic and onion have a demonstrated ability to inhibit platelet aggregation, possibly by interfering with prostaglandin biosynthesis (Ali et al. 1993).

In a double-blind, placebo-controlled study of 60 volunteers with cerebrovascular risk factors and constantly increased platelet aggregation, it was demonstrated that daily ingestion of 800 mg of powdered garlic (in the form of coated tablets), over four weeks, significantly decreased the ratio of circulating platelet aggregates and inhibited spontaneous platelet aggregation. The ratio of circulating platelet aggregates decreased by 10.3 percent; spontaneous platelet aggregation decreased by 56.3 percent (Kiesewetter et al. 1993).

Some garlic compounds that inhibit platelet aggregation have been identified. These are methyl (2-propenyl)trisulfide (the strongest), methyl (2-propenyl)disulfide, di(2-propenyl)disulfide, and di(2-propenyl)trisulfides. All these compounds are said to be formed from alliin, which is di(2-propenyl)thiolsulfinate. There is some evidence that methyl(2-propenyl)trisulfide is more effective on a molar basis than aspirin (Makheja, Vanderhoek, and Bailey 1979; Ariga, Oshiba, and Tamada 1981; Bosia et al. 1983; Apitz-Castro, Badimon, and Badimon 1992; Lawson, Ransom, and Hughes 1992). Recently, a novel amino acid glycoside, (-)-N-(1'-beta-D-fructopyranosyl)-S-2-propenyl-L-cysteine sulfoxide, showed significant inhibition of in vitro platelet aggregation induced by ADP (adenosin diphosphate) and epinephrine (Mutsch-Eckner et al. 1993).

Garlic has also been shown to increase fibrinolytic activity, which inhibits clot formation (Bordia et al. 1978). An excellent epidemiological study of garlic and onion intake in the Jain community in India was done in 1979. Three groups with widely differing allium consumption patterns were chosen: those who had always abstained from onions and garlic; those who consumed only small amounts (<200 g onion, <10 g garlic per week); and those who consumed

onions and garlic liberally (>600 g onion, >50 g garlic per week). The three groups were otherwise similar in regard to intake of calories, fat, and carbohydrates. Those who ingested the most alliums had the lowest level of plasma fibrinogen, which is used by the body in forming a blood clot with platelets (Sainani, Desai, Natsu et al. 1979).

In a study of dried garlic consumption by 20 patients with hyperlipoproteinemia over a period of four weeks, fibrinogen and fibrinopeptide A significantly decreased by 10 percent. Serum cholesterol levels significantly decreased by 10 percent. Systolic and diastolic blood pressure decreased. ADP- and collagen-induced platelet aggregation were not influenced (Harenberg, Giese, and Zimmermann 1988).

The antithrombotic agents found in garlic that we know about are (E,Z)-ajoene, or (E,Z)-4,5,9-trithiadodeca-1,6,11-triene 9-oxide, the major anticoagulant, di(2-propenyl)trisulfide, and 2-vinyl-4H-1,3-dithiene (Apitz-Castro et al. 1983; Block et al. 1984; Block 1992).

It is generally known that both fresh and boiled garlic decrease cholesterol and triglycerides. The Jain epidemiological study, mentioned previously, demonstrated not only that liberal use of onions and garlic decreased total cholesterol, low-density lipoprotein (LDL - the so-called bad cholesterol), and triglycerides, but also that those who consumed even small amounts of alliums were better protected than those who ate no onions or garlic (Sainani, Desai, Gorhe, et al. 1979).

One study, however, found that garlic increased cholesterol in people who had suffered a heart attack. A longer-term trial of garlic's effects on these people was undertaken and lasted 10 months. After 1 month, there was an increase in cholesterol, but thereafter it decreased, and after 8 months, it had declined by 18 percent. The initial increase in serum cholesterol in the heart patients who were fed garlic may have been caused by mobilization of lipid from deposits. Decreases of LDL occurred, and high-density lipoprotein (HDL - the "good" cholesterol) increased (Bordia 1981).

A multicentric, placebo-controlled, randomized study of standardized garlic-powder tablets in the treatment of hyperlipidemia (cholesterol levels over 200 mg/dl) was performed over a 16-week period. The total intake of garlic powder was 800 mg/day, standardized to 1.3 percent of alliin, (+)S-(2-propenyl)-L-cysteine, content (Stoll and Seebeck 1948). Cholesterol levels dropped 12 percent and triglyceride levels dropped 17 percent, with the best lowering effects seen in patients with cholesterol values between 250 to 300 mg/dl (Mader 1990).

To assess the effects of standardized garlic-powder tablets on serum lipids and lipoproteins, 42 healthy adults (19 men and 23 women), with a mean age of 52 (plus or minus 12 years), and with total serum cholesterol levels of 220 mg/dl or above, received, in a

randomized, double-blind fashion, 300 mg of standardized garlic powder (in tablet form) three times a day for 12 weeks, or they received a placebo. Diets and physical activities were unchanged. Treatment with standardized garlic at 900 mg/day produced a significantly greater reduction in serum triglycerides and LDL cholesterol than did the placebo. LDL-C (low-density lipoprotein cholesterol) was reduced 11 percent by garlic treatment and 3 percent by placebo ($p < 0.05$), and the baseline total cholesterol level of 262 (plus or minus 34 mg/dl) dropped to 247 (plus or minus 40 mg/dl) ($p < 0.01$). The placebo group showed a change from 276 (plus or minus 34 mg/dl) to 274 (plus or minus 29 mg/dl) (Jain et al. 1993).

Part of the activity of garlic results from an interruption of normal cholesterol biosynthesis (Qureshi et al. 1983). Hepatic cell culture results indicate that the hypocholesterolemic effect of garlic proceeds, in part, from decreased hepatic cholesterologenesis, whereas the triacylglycerol-lowering effect appears to be the result of the inhibition of fatty-acid synthesis (Yeh and Yeh 1994). The garlic compounds di(2-propenyl)thiosulfinate (allicin), S-methyl-L-cysteine sulfoxide, and S-(2-propenyl)-L-cysteine sulfoxide lower cholesterol in animals (Itokawa et al. 1973; Augusti and Matthew 1975). The garlic compounds ajoene, methylajoene, 2-vinyl-4H-1,3-dithiin, di(2-propenyl)disulfide, and allicin inhibit cholesterol synthesis in rat livers by 37 to 72 percent (Sendl et al. 1992). There is some evidence that di(2-propenyl)-disulfide inactivates 3-hydroxy-3-methylglutaryl-CoA (HMG-CoA) reductase (the major cholesterol synthesis enzyme) by forming an internal protein disulfide inaccessible for reduction and making the enzyme inactive (Omkumar et al. 1993). It has also been found that ajoene inactivates human gastric lipase, which causes less absorption of fat to occur in the digestion process and, therefore, lowers triacylglycerol levels (Gargouri et al. 1989).

Meta-analysis of the controlled trials of garlic's role in reducing hypercholesterolemia showed a significant reduction in total cholesterol levels. The best available evidence suggests that garlic, in an amount approximating one-half to one clove per day, decreased total serum cholesterol levels by about 9 percent in the groups of patients studied (Warshafsky, Kamer, and Sivak 1993; Silagy and Neil 1994).

Antimicrobial, antiviral, antihelminthic, and antifungal action. Garlic has been found to be more potent than 13 other spices in inhibiting *Shigella sonnei* (bacillary dysentery), *Staphylococcus aureus* (boils and food poisoning), *Escherichia coli* (indicator of fecal contamination), *Streptococcus faecalis* (indicator of fecal contamination), and *Lactobacillus casei* (found in milk and cheese) (Subrahmanyam, Krishnamurthy, et al. 1957; Subrahmanyam, Sreenivasamurthy, et al. 1957). A mouthwash containing 10 percent garlic extract has been shown to significantly reduce oral

bacterial counts (Elnima et al. 1983). However, the antibacterial components of garlic are heat labile. Whole garlic bulbs can lose their antibacterial activity within 20 minutes in boiling water at 100° C (Chen, Chang, and Chang 1985).

Garlic may change the composition of intestinal microflora to favor lactic organisms that are beneficial in the absorption of dietary minerals (Subrahmanyam, Krishnamurthy, et al. 1957; Subrahmanyam, Sreenivasamurthy, et al. 1957). Lactic acid bacteria have been proven to be the least sensitive microorganisms to the inhibitory effects of garlic.

In general, garlic is effective against most gram-positive and gram-negative bacteria. Garlic extracts inhibit the coagulase activity of *S. aureus* (Fletcher, Parker, and Hassett 1974). The *Listeria monocytogenes* population of strain Scott A (a source of food poisoning) was decreased to less than 10 per milliliter in seven days by 1 percent garlic powder (Hefnawy, Moustafa, and Marth 1993). Garlic also inhibits *Vibrio parahaemolyticus* (a gastroenteritis-causing pathogen in raw or improperly cooked fish or seafood) (Sato, Terao, and Ishibashi 1993).

Garlic has been found beneficial in cryptococcal meningitis, a frequently fatal disease (Fromtling and Bulmer 1978; Garlic in cryptococcal meningitis 1980; Caporaso, Smith, and Eng 1983). A commercial garlic extract, intravenously infused into two patients, caused their plasma titers of anti-*Cryptococcus neoformans* activity to rise twofold over preinfusion titers (Davis, Shen, and Cai 1990).

Thirty strains of mycobacteria, consisting of 17 species, were inhibited by various concentrations of garlic extract (1.34 to 3.35 mg/ml of agar media). Six strains of *Mycobacterium tuberculosis* required a mean inhibitory concentration of 1.67 mg/ml of media (Delaha and Garagusi 1985).

Garlic has proven effective in leprous neuritis. It is not certain whether this results from the vegetable's topical antibiotic activity or from garlic's ability to improve the thiamin status of the patient (Ramanujam 1962; Sreenivasamurthy et al. 1962).

Allicin, di(2-propenyl)thiosulfinate, the principal fresh-flavor component of garlic, is effective in the range of 1:125,000 against a number of gram-positive and gram-negative organisms. It inhibits the growth of some *Staphylococci*, *Streptococci*, *Vibrio* (including *Vibrio cholerae*), and *Bacilli* (including *Bacillus typhosus*, *Bacillus dysenteriae*, and *Bacillus enteritidis*) but is considerably weaker than penicillin against gram-positive organisms (Carson 1987). Allicin's effect is generally attributed to its interaction with biological -SH (sulfur) containing systems. If -SH-containing systems are necessary components for the growth and development of microorganisms, these processes will be inhibited by allicin. If the toxic compounds are exogenous, then reaction with allicin will lead to detoxification (Cavallito 1946; Wills 1956).

Garlic has exhibited antiviral activity against

influenza B virus and herpes simplex type I (nongenital) but not against Coxsaki B1 virus, which, however, usually causes only a mild illness (Carson 1987). Clinical use of garlic preparations in the prevention and treatment of human cytomegalovirus infections is effective (Meng et al. 1993). Because the antiviral effect of garlic extract is strongest when it is applied continuously in tissue culture, it is recommended that the clinical use of garlic extract against cytomegalovirus infection be persistent, and the prophylactic use of garlic extract is preferable in immunocompromised patients (Guo et al. 1993).

The activity of garlic constituents against selected viruses, including herpes simplex virus type 1 (nongenital cold sores), herpes simplex virus type 2 (genital), parainfluenza virus type 3 (bronchitis and pneumonia), vaccinia virus (cowpox, the source of an active vaccine against smallpox), vesicular stomatitis virus (which causes cold sores in humans and animals), and human rhinovirus type 2 (the common cold), has been determined. In general, the virucidal constituents, in descending order, were: ajoene, alliin, 2-propenyl methyl thiosulfinate, and methyl 2-propenyl thiosulfinate (Weber et al. 1992). Ajoene has also shown some activity against human immunodeficiency virus (HIV) (Tatarrintsev et al. 1992).

The effect of serial dilutions of crude garlic extract on adult *Hymenolepis nana* (dwarf tapeworm) was studied to detect the minimal lethal concentration. Garlic was then employed in the treatment of 10 children infected with *H. nana* and 26 children infected with *Giardia lamblia* (giardiasis). Such treatment took the form of either 5 milliliters of crude extract in 100 milliliters of water in two doses per day, or two commercially prepared 0.6 mg capsules twice a day for three days. Garlic was found to be efficient and safe and to shorten the duration of treatment (Soffar and Mokhtar 1991). Garlic appears to affect adversely the development of the eggs of *Necator americanus* (hookworm) but has less effect on the hatched larvae (Bastidas 1969). Rectal garlic preparations may be effective in the treatment of pinworms (Braun 1974).

A single dose of ajoene on the day of malarial infection was found to suppress the development of parasitemia; there were no obvious acute toxic effects from the tested dose. The combination of ajoene and chloroquine, given as a single dose on the day of the infection, completely prevented the subsequent development of malarial parasitemia in treated mice (Perez, de la Rosa, and Apitz 1994).

Ajoene has also been shown to inhibit the proliferation of *Trypanosoma cruzi*, the causative agent of Chagas' disease. An important factor associated with the antiproliferative effects of ajoene against *T. cruzi* may be its specific alteration of the phospholipid composition of these cells (Urbina 1993).

Garlic inhibits the aflatoxin-producing fungi *Aspergillus flavus* and *Aspergillus parasiticus*

(Sharma et al. 1979). Garlic extract inhibits the growth and aflatoxin production of *A. flavus* (Sutabaha, Suttajt, and Niyomca 1992), and garlic oil completely inhibits sterigmatocystin (a carcinogenic mycotoxin produced by *Aspergillus*) production (Hasan and Mahmoud 1993). Thiopropanal-S-oxide is one of the most active antiaflatoxin components (Sharma et al. 1979).

The ajoene in garlic has been shown to have antifungal activity. *Aspergillus niger* (a frequent cause of fungal ear infections) and *Candida albicans* (yeast) were inhibited by ajoene in concentrations of less than 20 micrograms per milliliter (Yoshida et al. 1987). Ajoene also inhibits the growth of the pathogenic fungus *Paracoccidioides brasiliensis* (South American blastomycosis, which starts in the lungs) (San Blas et al. 1993). Additional studies have shown ajoene to inhibit *Cladosporium carrionii* and *Fonsecaea pedrosoi* (both cause chromoblastomycosis, a fungal disease of the skin) (Sanchez-Mirt, Gil, and Apitz-Castro 1993).

Moreover, extracts of both garlic and onion have been shown to inhibit the growth of many plant-pathogenic fungi and yeasts. Garlic-bulb extracts are more active than onion extracts (Agrawal 1978). Garlic solutions of 1 to 20 percent have been effective against plant pathogens such as downy mildew in cucumbers and radishes, bean rust, bean anthracnose, tomato early blight, brown rot in stone fruits, angular leaf spot in cucumbers, and bacterial blight in beans (Pordesimo and Ilag 1976). Ajoene has been tested in greenhouse experiments, where it completely inhibited powdery mildew in tomatoes and roses (Reimers et al. 1993).

Anticarcinogenic. Some data have suggested an inverse relationship between garlic consumption and gastric cancer. In Shandong Province, China, the death rate from gastric cancer was found to be 3.45/100,000 population in Gangshan County (where garlic consumption is approximately 20 g per person per day), but in nearby Quixia County (where little garlic is eaten), the gastric cancer death rate was much higher, averaging 40/100,000 (Han 1993; Witte et al. 1996). A study of risk factors for colon cancer in Shanghai indicated that garlic was associated with a decreased relative risk (Yang, Ji, and Gao 1993). Some evidence to the same effect has been seen in Italy (Dorant et al. 1993).

Interviews with 564 patients with stomach cancer and 1,131 controls – in an area of China where gastric cancer rates were high – revealed a significant reduction in gastric cancer risk with increasing consumption of allium vegetables. Persons in the highest quartile of intake experienced only 40 percent of the risk of those in the lowest quartile. Protective effects were seen for garlic, onions, and other allium foods. Although additional research is needed before etiologic inferences can be made, the findings were con-

sistent with reports of tumor inhibition following administration of allium compounds in experimental animals (You et al. 1989). Garlic has been shown to reduce cancer promotion and tumor yield by phorbol-myristate-acetate in mice (Belman 1983).

In isolated epidermal cells, at 5 µg per milliliter, garlic oil increased glutathione peroxidase activity and inhibited ornithine decarboxylase induction in the presence of various nonphorbol ester tumor promoters. The same oil treatment inhibited the sharp decline in the intracellular ratio of reduced glutathione to oxidized glutathione caused by the potent tumor promoter, 12-O-tetradecanoylphorbol-13-acetate. It was suggested that some of the inhibitory effects of garlic on skin tumor promotion may have resulted from its enhancement of the natural glutathione-dependent antioxidant protective system of the epidermal cells (Perchellet et al. 1986). The active compound appeared to be di(2-propenyl)trisulfide (Carson 1987).

Other medicinal uses. Garlic has been used to treat hypertension in China and Japan for centuries. Studies in 1921, 1948, and 1969 provided supporting evidence of garlic's antihypertensive ability (Loeper and Debray 1921; Piotrowski 1948; Srinivasan 1969).

In 1990, a study was published in which 47 outpatients with mild hypertension took part in a randomized, placebo-controlled, double-blind trial conducted by 11 general practitioners. The patients who were admitted to the study had diastolic blood pressures between 95 and 104 mm Hg. The patients took either a preparation of garlic powder or a placebo of identical appearance for 12 weeks. Blood pressure and plasma lipids were monitored during treatment at 4, 8, and 12 weeks. Significant differences between the placebo and garlic groups were found during the course of therapy. The supine diastolic blood pressure in the group taking garlic fell from 102 to 91 mm Hg after 8 weeks ($p < 0.05$) and to 89 mm Hg after 12 weeks ($p < 0.01$). Serum cholesterol and triglycerides were also significantly reduced after 8 and 12 weeks of treatment. In the placebo group no significant changes occurred (Auer et al. 1990).

Studies of natural selenium-rich sources have found that high-selenium garlic and onion may have some unique attributes. First, their ingestion does not lead to an exaggerated accumulation of tissue selenium, which both selenomethionine and Brazil nut may cause. Second, unlike selenite, they do not cause any perturbation in glutathione (an antioxidant) homeostasis. Third, they expressed good anticancer activity that was equal to, if not better than, that of selenite (Ip and Lisk 1994).

Garlic odor. Although the problem of onion and garlic breath was first investigated in 1935 (Haggard and Greenberg 1935), many folk remedies – such as strong coffee, honey, yogurt, milk, coffee beans,

cloves, and, most commonly, parsley – have long been used (Sokolov 1975). Perhaps, however, there is excessive worry about garlic or onion on the breath. A recent study of male and female shoppers in Helsinki indicated that sweat and alcohol were thought to be the most annoying social odors and those of garlic and perfume or aftershave the least annoying (Rosin, Tuorila, and Uutela 1992).

Studies on the effect of garlic on breast milk have indicated that garlic ingestion significantly and consistently increases the intensity of the milk odor. It was found that infants were attached to the breast for longer periods of time and sucked more when the milk smelled of garlic. There was also a tendency for the infants to ingest more milk. However, if the mother ingested garlic pills regularly, there was no change in the infant's feeding behavior after its initial exposure (Mennella and Beauchamp 1991, 1993).

Leeks

History

As with onions and garlic, leek (*Allium porrum*) consumption is depicted in Egyptian tomb decorations of the Early Dynastic Period (c. 2925 B.C.–c. 2575 B.C.) (Jones and Mann 1963; Fenwick and Hanley 1985a). Leeks were also grown and widely used for cooking in Sumeria (southern Iraq) even earlier (Hanley and Fenwick 1985). In China, the fifth-century treatise on agriculture, *Ch'i-min-yao-shu* (Essential Arts for the People), by Chia Ssu-hsieh described the cultivation of *ch'iu* (Chinese leek, *Allium ramosum* L.) along the Red River valley, where it has doubtless been cultivated for many centuries (Li 1969).

Leeks were called *prason* in ancient Greece and *porrum* by the Romans. Pliny the Elder, the Roman naturalist, cited Aricia in central Italy as famous for its leeks, and the Emperor Nero (A.D. 37–68) reportedly ate them several days a month to clear his voice, which caused people to call him Porrophagus. The Romans introduced leeks to Britain, where they were widely cultivated by Saxon times (sixth century A.D.), and cottage vegetable plots were often referred to by the name “leac tun” (Hedrick 1972; Fenwick and Hanley 1985a).

Leeks were known in Europe throughout the Middle Ages and were believed – like onions and garlic – to be an erotic stimulant that increased sperm and stimulated desire, especially when prepared with honey, sesame, and almond (Mauron 1986).

In northern England, leek growing remains a serious and highly competitive business, with secrets of cultivation handed down from father to son. In addition, leeks have been the badge of Welshmen from time immemorial. Saint David (c. 495–589) is said to have suggested that the Welsh wear leeks in their hats to enable them to distinguish friend from foe in the heat of battle. Consequently, the leek is worn (and subsequently eaten) in Wales on St. David's Day

(March 1) to celebrate the Welsh defeat of the Saxons in the year 633 (Hedrick 1972; Fenwick and Hanley 1985a).

Horticulture and Botany

The modern leek is not known in the wild. It probably originated in the Near East region around the eastern Mediterranean, where it was much eaten, and was distributed across Europe by the Romans (Traub 1968).

Cultivation. Although leek growing is popular in parts of Britain, commercial production of the plant is centered in France, Belgium, and the Netherlands, with France by far the most important grower (Hanley and Fenwick 1985). Production takes place mainly in Bouches-du-Rhône, Vaucluse, Haute Garonne, Ain, Ille et Vilaine, Manche, and especially in Nord and Loire-Atlantique.

Leeks grow well under most soil conditions but do best in deep loams and peat. Good drainage is essential, and the soil's Ph value should be near 7.0. Leeks can be sowed directly or grown in seedbeds and transplanted. Six varieties of leeks are grown in Britain to ensure year-round cultivation. Harvesting may be mechanical or by hand. A maximum yield of fresh weight and dry matter can be obtained after harvest in October or November (weeks 43 to 45), when nitrate content has decreased to a low and almost stable level (Kaack, Kjeldsen, and Mune 1993). Leeks are then trimmed, either in the field or at a packing station (Fenwick and Hanley 1985a). Leeks store well at 0° C (with 90 to 95 percent relative humidity) for up to 12 weeks (Vandenberg and Lentz 1974).

Nutrition

The Nutrition Data System of the University of Minnesota indicates that 100 g of leeks provides 32 kilocalories of energy, 1.83 g of protein, 0.19 g of fat, and 7.34 g of carbohydrate (Nutrition Coordinating Center 1994).

Approximately one-half cup of fresh leeks would give an 18- to 25-year-old male 9 percent of his RDA of calcium (72 mg), 14.8 percent of iron (1.48 mg), 31.3 percent of vitamin C (18.8 mg), and 32 percent of his folacin (64 mg). Leeks are high in potassium (276 mg/100 g), low in sodium (16 mg/100 g), and contain small amounts of copper, magnesium, phosphorus, selenium, and zinc. They have 38.42 mcg of vitamin A, 230.54 mcg of beta-carotene, and 0.46 mg of vitamin E (alpha-tocopherol 0.37 mg, beta-tocopherol 0.17 mg, gamma-tocopherol 0.17 mg, and delta-tocopherol 0.09 mg), as well as small amounts of thiamine, riboflavin, niacin, pantothenic acid, and vitamin B₆ (National Research Council 1989). All essential amino acids are present in leeks. Aspartic acid (0.17 g) and glutamic acid (0.38 g) are the most abundant, followed by arginine (0.13 g) and proline (0.12 g).

Chemistry

Nonflavor compounds. The flavonoids most often found in leeks have been quercetin, kaempferol, and their derivatives, usually mono- and diglycosides. These are generally found in higher concentrations in the epidermal layer of the leaves and protect the plant from ultraviolet radiation (Starke and Herrmann 1976b).

Leeks contain fructans, polymers of fructose usually having 3 to 12 fructose molecules. Fructose polymers of 12 molecules are the most common (Darbyshire and Henry 1981). Leeks also produce very long chain fatty acids (Agrawal, Lessire, and Stumpf 1984).

Chives

History

Chives (*Allium schoenoprasum*) originated in the north temperate zone. John Gerard (1545–1612), English botanist and barber-surgeon, included chives in his herbal, published in 1597. Described in 1683 as a pleasant sauce and food potherb, and listed as part of seeds-men's supplies in 1726, chives were losing favor in England by 1783. However, botanist E. Louis Sturtevant reported in the nineteenth century that Scottish families were still heavy chive consumers (Hedrick 1972).

Chives are cultivated for use in salads and soups, and many consider them an indispensable ingredient in omelets. They have been much used for flavoring in continental Europe, especially in Catholic countries. Chives were also included in an 1806 list of American esculents (Hedrick 1972).

Horticulture and Botany

Chives are the only one of the allium species native to both the Old World and the New (Simmonds 1976). Indeed, the plant's wild form occurs in Asia as well as in North America and Europe.

Chives flower in spring and summer, and bees are important for their fertilization (Nordestgaard 1983). The plants grow in dense clumps of narrow cylindrical leaves and taller hollow flower stems with globular heads. Their bulbs are elongated and only slightly swollen, but it is the leaves that are usually chopped and used as a garnish for other foods. The plant is mostly homegrown and is also used as an ornamental (Traub 1968).

Pathogens and Pests

As with onions, chives are subject to assault from downy mildew (*P. destructor* [Berk.] Casp.) and onion smut (*U. cepulae* Frost), as well as the bulb and stem nematode (*D. dispa* [Kuhn] Filipjcr).

Food Use and Nutrition

Chives are eaten fresh or dehydrated, the latter being the most common processed form today. The flavor of the chopped leaves remains stable for several months when deep-frozen or freeze-dried (Poulsen and Nielsen 1979).

As with the other alliums, the nutritional content of chives varies by variety, ecological conditions, and climate. One hundred grams of chives will generally provide about 30 kilocalories of energy, 3.27 g of protein, 0.73 g of fat, and 4.35 g of carbohydrate (Nutrition Coordinating Center 1994).

For a male between 18 and 25 years of age, approximately one-half cup of fresh chives delivers 11.5 percent of the RDA of calcium (92 mg), 16 percent of iron (1.6 mg), 12 percent of magnesium (42 mg), 43.4 percent of vitamin A (434.43 mcg RE), 96.8 percent of vitamin C (58.1 mg), and 52.5 percent of folacin (105 mg). Chives are high in potassium (296 mg) and low in sodium (3 mg). They contain small amounts of copper, phosphorus, and zinc. Chives have 2606.59 mcg of beta-carotene and 0.46 mg of vitamin E (alpha-tocopherol 0.37 mg, beta-tocopherol 0.17 mg, gamma-tocopherol 0.17 mg, and delta-tocopherol 0.09 mg) and small amounts of thiamine, riboflavin, niacin, pantothenic acid, and vitamin B₆ (National Research Council 1989).

Medicinal Use

Chives have some antibacterial effects (Huddleson et al. 1944). Extracts of onions and chives possess tuberculostatic activity against human, avian, and bovine strains. In fact, chives show rather more activity than onions and are only slightly less effective than streptomycin (Gupta and Viswanathan 1955). In addition, aqueous extracts of chives have exhibited significant activity against leukemia in mice (Caldes and Prescott 1973).

Shallots

History

Pliny the Elder, in his *Natural History*, mentioned the Ascalon onion (the shallot, *Allium ascalonicum*) as one of six types of onions known to the Greeks (Fenwick and Hanley 1985a). He wrote that it came from Ascalon in Syria, and Joseph Michaud's history of the Crusades affirmed this origin. Shallots were known in Spain, Italy, France, and Germany by 1554, had entered England from France by 1633, and were grown in American gardens by 1806 (Hedrick 1972).

Horticulture and Botany

Shallots were once viewed as a separate species, but botanists now consider them to be a variety of *A. cepa* L. They are cultivated ubiquitously (Hedrick 1972) but not extensively, save in the Netherlands and France (Fenwick and Hanley 1985a).

Food Use and Nutrition

Shallots can be dried in the field, weather permitting. They are employed as a seasoning in stews and soups but can also be used in the raw state, diced in salads, or sprinkled over steaks and chops. Shallots also make excellent pickles (Hedrick 1972).

As with the rest of the alliums, the nutritional content of shallots depends on variety, ecological conditions, and climate. According to the Nutrition Coordinating Center (1994), 100 g (3.53 ounces or 0.44 cup) of shallots yields 32 kilocalories of energy, 1.83 g of protein, 0.19 g of fat, and 7.34 g of carbohydrate.

One-half cup of fresh shallots provides a male between 18 and 25 years of age approximately 9 percent of the RDA of calcium (72 mg), 14.8 percent of iron (1.48 mg), 31.3 percent of vitamin C (18.8 mg), and 32 percent of folacin (64 mg) (National Research Council 1989). Shallots are high in potassium (276 mg) and low in sodium (16 mg). They contain small amounts of copper, magnesium, phosphorus, selenium, and zinc, and also have 38.42 mcg RE of vitamin A (230.54 mcg of beta-carotene) and 0.46 mg of vitamin E (alpha-tocopherol 0.37 mg, beta-tocopherol 0.17 mg, gamma-tocopherol 0.17 mg, and delta-tocopherol 0.09 mg), as well as small amounts of thiamine, riboflavin, niacin, pantothenic acid, and vitamin B₆.

Flavor Compounds

Shallots have the same flavor components as onions but generally contain more methyl, propyl, and (1-propenyl) di- and trisulfides (Dembele and Dubois 1973; Wu et al. 1982).

A study of the volatile oils from raw, baked, and deep-fried shallots identified sulfides, disulfides, trisulfides, thiophene derivatives, and oxygenated compounds. The oils from baked or fried shallots contain decreased amounts of alkyl propenyl disulfides and increased amounts of dimethyl thiophenes (Carson 1987).

Julia Peterson

Bibliography

- Adamicki, F., and A. K. Kepka. 1974. Storage of onions in controlled atmospheres. *Acta Horticultura* 38: 53.
- Adetumbi, M., G. T. Javor, and B. H. Lau. 1986. *Allium sativum* (garlic) inhibits lipid synthesis by *Candida albicans*. *Antimicrobial Agents and Chemotherapy* 30: 499-501.
- Agrawal, P. 1978. Effect of root and bulb extracts of *Allium* spp. on fungal growth. *Transactions of the British Mycological Society* 70: 439.
- Agrawal, R. K., H. A. Dewar, D. J. Newell, and B. Das. 1977. Controlled trial of the effect of cycloalliin on the fibrinolytic activity of venous blood. *Atherosclerosis* 27: 347.
- Agrawal, V. P., R. Lessire, and P. K. Stumpf. 1984. Biosynthesis of very long chain fatty acids in microsomes from epidermal cells of *Allium porrum* L. *Archives of Biochemistry and Biophysics* 230: 580.
- Alexander, M. M., and G. A. Sulebele. 1973. Pectic substances in onion and garlic skins. *Journal of the Science of Food and Agriculture* 24: 611.
- Ali, M., M. Angelo Khattar, A. Parid, et al. 1993. Aqueous extracts of garlic (*Allium sativum*) inhibit prostaglandin synthesis in the ovine ureter. *Prostaglandins Leukotrienes and Essential Fatty Acids* 49: 855-9.
- Allen, M. L., M. H. Mellow, M. G. Robinson, and W. C. Orr. 1990. The effect of raw onions on acid reflux and reflux symptoms. *American Journal of Gastroenterology* 85: 377-80.
- Amla, V., S. L. Verma, T. R. Sharma, et al. 1980. Clinical study of *Allium cepa* Linn in patients of bronchial asthma. *Indian Journal of Pharmacology* 13: 63.
- Amonkar, S. V., and A. Banerji. 1971. Isolation and characterization of larvicidal principle of garlic. *Science* 174: 1343.
- Anesini, C., and C. Perez. 1993. Screening of plants used in Argentine folk medicine for antimicrobial activity. *Journal of Ethnopharmacology* 39: 119-28.
- Apitz-Castro, R., J. J. Badimon, and L. Badimon. 1992. Effect of ajoene, the major antiplatelet compound from garlic, on platelet thrombus formation. *Thrombosis Research* 68: 145-55.
- Apitz-Castro, R., S. Cabrera, M. R. Cruz, et al. 1983. The effects of garlic extract and of three pure components isolated from it on human platelet aggregation, arachidonate metabolism, release activity and platelet ultrastructure. *Thrombosis Research* 32: 155.
- Ariga, T., S. Oshiba, and T. Tamada. 1981. Platelet aggregation inhibitor in garlic. *Lancet* 8212: 150.
- Auer, W., A. Eiber, E. Hertkorn, et al. 1990. Hypertension and hyperlipidaemia: Garlic helps in mild cases. *British Journal of Clinical Practice - Symposium Supplement* 69: 3-6.
- Augusti, K. T. 1974. Effect on alloxan diabetes of allyl propyl disulphide obtained from onion. *Die Naturwissenschaften* 61: 172.
- Augusti, K. T., and M. E. Benaim. 1975. Effect of essential oil of onion (allyl propyl disulphide) on blood glucose, free fatty acid and insulin levels of normal subjects. *Clinica Chimica Acta* 60: 121.
- Augusti, K. T., and P. T. Matthew. 1975. Effect of allicin on certain enzymes of liver after a short-term feeding to normal rats. *Experientia* 31: 148.
- Auro de Ocampo, A., and E. M. Jimenez. 1993. Plant medicine in the treatment of fish diseases in Mexico. *Veterinaria-Mexico* 24: 291-5.
- Austin, R. B. 1972. Bulb formation in onions as affected by photoperiod and spectral quality of light. *Journal of Horticultural Science* 47: 493.
- Barone, F. E., and M. R. Tansey. 1977. Isolation, purification, identification, synthesis, and kinetics of activity of the anticandidal component of *Allium sativum*, and a hypothesis for its mode of action. *Mycologia* 69: 793.
- Bartzatt, R., D. Blum, and D. Nagel. 1992. Isolation of garlic derived sulfur compounds from urine. *Analytical Letters* 25: 1217-24.
- Bastidas, G. J. 1969. Effect of ingested garlic on *Necator americanus* and *Ancylostoma caninum*. *American Journal of Tropical Medicine and Hygiene* 18: 920-3.
- Belman, S. 1983. Onion and garlic oils inhibit tumour promotion. *Carcinogenesis* 4: 1063.
- Bezanger-Beauquesne, L., and A. Delelis. 1967. Sur les flavonoides du bulbe d'*Allium ascalonicum* (Liliaceae). *Compte Rendu Académie Scientifique Paris Series D* 265: 2118.
- Bhatnagar-Thomas, P. L., and A. K. Pal. 1974. Studies on the insecticidal activity of garlic oil. II. Mode of action of the oil as a pesticide in *Musca domestica* nebulosa Fabr and *Trogoderma granarium* Everts. *Journal of Food Science and Technology (Mysore)* 11: 153.

- Bhushan, S., S. Verma, V. M. Bhatnagar, and J. B. Singh. 1976. A study of the hypocholesterolaemic effect of onion (*Allium cepa*) on normal human beings. *Indian Journal of Physiology and Pharmacology* 20: 107.
- Bierman, C. J. 1983. Insect repellent. Belgian Patent BE 896,522 (Cl.AOIN), August 16. NL Application 82/2, 260, June 4, 1982.
- Bingham, S., J. H. Cummings, and N. I. McNeil. 1982. Diet and health of people with an ileostomy. 1. Dietary assessment. *British Journal of Nutrition* 47: 399-406.
- Bioelens, M., P. J. deValois, H. J. Wobben, and A. van der Gern. 1971. Volatile flavor compounds from onion. *Journal of Agriculture and Food Chemistry* 19: 984.
- Block, E. 1992. The organosulfur chemistry of the genus *Allium*: Implications for the organic chemistry of sulfur. *Angewandte Chemie International Edition in English* 31: 1135-78.
- Block, E., S. Ahmad, M. K. Jain, et al. 1984. (E-Z)-ajoene - a potent antithrombic agent from garlic. *Journal of the American Chemical Society* 106: 8295.
- Block, E., P. E. Penn, and L. K. Revelle. 1979. Structure and origin of the onion lachrymatory factor. A microwave study. *Journal of the American Chemical Society* 101: 2200.
- Booth, S., T. Johns, and C. Y. Lopez-Palacios. 1993. Factors influencing self-diagnosis and treatment of perceived helminthic infection in a rural Guatemalan community. *Social Science and Medicine* 37: 531-9.
- Bordia, A. 1981. Effect of garlic on blood lipids in patients with coronary heart disease. *American Journal of Clinical Nutrition* 34: 2100.
- Bordia, A. K., S. K. Sanadhya, A. S. Rathore, et al. 1978. Essential oil of garlic on blood lipids and fibrinolytic activity in patients with coronary artery disease. *Journal of the Association of Physicians of India* 26: 327.
- Bosia, A., P. Spangenberg, W. Losche, et al. 1983. The role of the GSH-disulphide status in the reversible and irreversible aggregation of human platelets. *Thrombosis Research* 30: 137.
- Braun, H. 1974. *Heilpflanzen - Lexikon für Ärzte und Apotheker*. Stuttgart.
- Brewster, J. L. 1977a. The physiology of the onion. I. *Horticultural Abstracts* 47: 17.
- 1977b. The physiology of the onion. II. *Horticultural Abstracts* 47: 103.
- Brocklehurst, T. F., C. A. White, and C. Dennis. 1983. The microflora of stored coleslaw and factors affecting the growth of spoilage yeasts in coleslaw. *Journal of Applied Bacteriology* 55: 57.
- Brodnitz, M. H., and J. V. Pascale. 1971. Thiopropanal-S-oxide, a lachrymatory factor in onions. *Journal of Agriculture and Food Chemistry* 19: 269.
- Brodnitz, M. H., J. V. Pascale, and L. Vanderslice. 1971. Flavour components of garlic extract. *Journal of Agriculture and Food Chemistry* 19: 273.
- Brodnitz, M. H., C. L. Pollock, and P. P. Vallon. 1969. Flavour components of onion oil. *Journal of Agriculture and Food Chemistry* 17: 760.
- Buffington, D. E., S. K. Sastry, J. C. Gustashaw, Jr., and D. S. Burgis. 1981. Artificial curing and storage of Florida onions. *Transactions of the American Society of Agricultural Engineers* 2: 782.
- Caldes, G., and B. Prescott. 1973. A potential antileukemic substance present in *Allium ascalonicum*. *Planta Medica* 23: 99.
- Caporaso, N., S. M. Smith, and R. H. K. Eng. 1983. Antifungal activity in human urine and serum after ingestion of garlic (*Allium sativum*). *Antimicrobial Agents and Chemotherapy* 23: 700.
- Carson, J. F. 1987. Chemistry and biological properties of onions and garlic. *Food Reviews International* 3: 71-103.
- Cavallito, C. J. 1946. Relationship of thiol structures to reaction with antibiotics. *Journal of Biological Chemistry* 164: 29.
- Chen, H. C., M. D. Chang, and T. J. Chang. 1985. Antibacterial properties of some spice plants before and after heat treatment. *Chinese Journal of Microbiology and Immunology* 18: 190-5.
- Curzio, O. A., and C. A. Croci. 1983. Extending onion storage life by gamma-irradiation. *Journal of Food Processing and Preservation* 7: 19.
- Darbyshire, B., and R. J. Henry. 1978. The distribution of fructans in onions. *New Phytologist* 81: 29.
1981. Differences in fructan content and synthesis in some *Allium* species. *New Phytologist* 87: 249.
- Davis, L. E., J. K. Shen, and Y. Cai. 1990. Antifungal activity in human cerebrospinal fluid and plasma after intravenous administration of *Allium sativum*. *Antimicrobial Agents and Chemotherapy* 34: 651-3.
- Deb-Kirtaniya, S., M. R. Ghosh, N. Adityachaudhury, and A. Chatterjee. 1980. Extracts of garlic as possible source of insecticides. *Indian Journal of Agricultural Science* 50: 507.
- Delaha, E. C., and V. F. Garagusi. 1985. Inhibition of mycobacteria by garlic extract (*Allium sativum*). *Antimicrobial Agents and Chemotherapy* 27: 485-6.
- Dembele, S., and P. Dubois. 1973. Composition d'essence shallots (*Allium cepa* L. var. *aggregatum*). *Annales de Technologie Agricole* 22: 121.
- DeWit, J. C., S. Notermans, N. Gorin, and E. H. Kampelmacher. 1979. Effect of garlic oil or onion oil on toxin production by *Clostridium botulinum* in meat slurry. *Journal of Food Protection* 42: 222.
- Domiciano, N. L., A. Y. Ota, and C. R. Tedardi. 1993. Proper time for chemical control of thrips-Tabaci Lindeman 1888 on onion allium-cepal. *Anais da Sociedade Entomologica do Brasil* 2: 71-6.
- Dorant, E., P. A. van den Brandt, R. A. Goldbohm, et al. 1993. Garlic and its significance for the prevention of cancer in humans: A critical view. *British Journal of Cancer* 67: 424-9.
- Dorsch, W., O. Adam, J. Weber, and T. Ziegeltrum. 1985. Antiasthmatic effects of onion extracts - detection of benzyl - and other isothiocyanates (mustard oils) as antiasthmatic compounds of plant origin. *European Journal of Pharmacology* 107: 17.
- Ellmore, G. S., and R. S. Feldberg. 1994. Alliin lyase localization in bundle sheaths of the garlic clove (*Allium sativum*). *American Journal of Botany* 81: 89-94.
- Elnima, E. I., S. A. Ahmed, A. G. Mekki, and J. S. Mossa. 1983. The antimicrobial activity of garlic and onion extracts. *Pharmazie* 38: 747.
- Falleroni, A. E., C. R. Zeiss, and D. Levitz. 1981. Occupational asthma secondary to inhalation of garlic dust. *Journal of Allergy and Clinical Immunology* 68: 156.
- Farkas, G. L., and Z. Kiraly. 1962. Role of phenolic compounds on the physiology of plant diseases and disease resistance. *Phytopathologie Zeitschrift* 44: 105.
- Farkas, M. C., and J. N. Farkas. 1974. Hemolytic anemia due to ingestion of onions in a dog. *Journal of the American Animal Hospital Association* 10: 65.
- Fenwick, G. R., and A. B. Hanley. 1985a. The genus *Allium*.

- Part 1. *Critical Reviews in Food Science and Nutrition* 22: 199-271.
- 1985b. The genus *Allium*. Part 2. *Critical Reviews in Food Science and Nutrition* 22: 273-377.
- 1985c. The genus *Allium*. Part 3. *Critical Reviews in Food Science and Nutrition* 23: 1-73.
- Fiskesjo, G. 1988. The *Allium* test - an alternative in environmental studies: The relative toxicity of metal ions. *Mutation Research* 197: 243-60.
- Fletcher, R. D., B. Parker, and M. Hassett. 1974. Inhibition of coagulase activity and growth of *Staphylococcus aureus* by garlic extracts. *Folia Microbiologica* 19: 494.
- FAO (Food and Agriculture Organization of the United Nations). 1992a. *FAO Yearbook Production* 46: 151-4.
- 1992b. *FAO Yearbook Trade* 46: 125-7.
- Freeman, G. G. 1975. Distribution of flavor components in onion (*Allium cepa* L.), leek (*Allium porrum*) and garlic (*Allium sativum*). *Journal of the Science of Food and Agriculture* 26: 471.
- Fromtling, R. A., and G. A. Bulmer. 1978. In vitro effect of aqueous extract of garlic (*Allium sativum*) on the growth and viability of *Cryptococcus neoformans*. *Mycologia* 70: 397.
- Fujiwara, M. 1976. Allithiamine and its properties. *Journal of Nutritional Science and Vitaminology* 22: 57.
- Fuleki, T. 1971. Anthocyanins in red onions, *Allium cepa*. *Journal of Food Science* 36: 101.
- Gargouri, Y., H. Moreau, M. K. Jain, et al. 1989. Ajoene prevents fat digestion by human gastric lipase in vitro. *Biochimica et Biophysica Acta* 1006: 137-9.
- Garlic in cryptococcal meningitis - a preliminary report of 21 cases. 1980. *Chinese Medical Journal* 93: 123.
- Giron, L. M., G. A. Aguilar, A. Caceres, et al. 1988. Anticandidal activity of plants used for the treatment of vaginitis in Guatemala and clinical trial of a *Solanum nigrescens* preparation. *Journal of Ethnopharmacology* 22: 307-13.
- Goldsmith, W. W. 1909. Onion poisoning in cattle. *Journal of Comparative Pathology and Therapy* 22: 151.
- Goodenough, P. W., and R. K. Atkin. 1981. *Quality in stored and processed vegetables and fruit*. New York.
- Graham, S., B. Haughey, J. Marshall, et al. 1990. Diet in the epidemiology of gastric cancer. *Nutrition and Cancer* 13: 19-34.
- Granroth, B. 1970. Biosynthesis and decomposition of cysteine derivatives in onion and other *Allium* species. *Annales Academiæ Scientiarum Fennicæ* 154 Series A. II. Chemica: 9.
- Granroth, B., and A. I. Virtanen. 1967. S-(2-carboxypropyl) cysteine and its sulphoxide as precursors in the biosynthesis of cycloalliin. *Acta Chemica Scandinavica* 21: 1654.
- Greco, N. 1993. Epidemiology and management of *Ditylenchus dipsaci* on vegetable crops in southern Italy. *Nematropica* 23: 247-51.
- Gruhzit, O. M. 1931. Anemia in dogs produced by feeding disulphide compounds. Part II. *American Journal of Medical Sciences* 181: 815.
- Gummary, C. S. 1977. A review of commercial onion products. *Food Trade Review* 47: 452.
- Guo, N. L., D. P. Lu, G. L. Woods, et al. 1993. Demonstration of the anti-viral activity of garlic extract against human cytomegalovirus in vitro. *Chinese Medical Journal* 106: 93-6.
- Gupta, K. C., and R. Viswanathan. 1955. In vitro study of anti-tubercular substances from *Allium* species. I. *Allium schoenoprasum*. II. *Allium cepa*. *Antibiotics and Chemotherapy* 5: 18.
- Haggard, H. W., and L. A. Greenberg. 1935. Breath odours from alliaceous substances. *Journal of the American Medical Association* 104: 2160.
- Hall, C. W. 1980. *Drying and storage of agricultural crops*. Westport, Conn.
- Han, J. 1993. Highlights of the cancer chemoprevention studies in China. *Preventive Medicine* 22: 712-22.
- Handa, G., J. Singh, and C. K. Atal. 1983. Antiasthmatic principle of *Allium cepa* Linn (onions). *Indian Drugs* 20: 239.
- Hanley, A. B., and G. R. Fenwick. 1985. Cultivated alliums. *Journal of Plant Foods* 6: 211-38.
- Hanson, L. P. 1975. *Commercial processing of vegetables*. Park Ridge, N.J.
- Harenberg, J., C. Giese, and R. Zimmermann. 1988. Effect of dried garlic on blood coagulation, fibrinolysis, platelet aggregation and serum cholesterol levels in patients with hyperlipoproteinemia. *Atherosclerosis* 74: 247-9.
- Hasan, H. A., and A. L. Mahmoud. 1993. Inhibitory effect of spice oils on lipase and mycotoxin production. *Zentralblatt für Mikrobiologie* 148: 543-8.
- Hashimoto, S., M. Miyazawa, and H. Kameoka. 1983. Volatile flavor components of chive (*Allium schoenoprasum* L.). *Journal of Food Science* 48: 1858.
- Heath, H. B. 1981. *Source book on flavors*. Westport, Conn.
- Hedrick, U. P. 1972. *Sturtevant's edible plants of the world*. New York.
- Hefnawy, Y. A., S. I. Moustafa, and E. H. Marth. 1993. Sensitivity of *Listeria monocytogenes* to selected spices. *Journal of Food Protection* 56: 876-8.
- Henson, G. E. 1940. Garlic, an occupational factor in the etiology of bronchial asthma. *Journal of the Florida Medical Association* 27: 86.
- Herrmann, K. 1958. Flavonols and phenols of the onion (*Allium cepa*). *Archiv der Pharmazie* 291: 238.
- Huddleson, I. F., J. Dufrein, K. C. Barrons, and M. Giefel. 1944. Antibacterial substances in plants. *Journal of the American Veterinary Medical Association* 105: 394.
- Hyams, E. 1971. *Plants in the service of man; 10,000 years of domestication*. London.
- IOS (International Organization for Standardization). 1983. *Garlic guide to cold storage*. International Standard ISO, New York.
- Ip, C., and D. J. Lisk. 1994. Characterization of tissue selenium profiles and anticarcinogenic responses in rats fed natural sources of selenium-rich products. *Carcinogenesis* 15: 573-6.
- Itoh, T., T. Tamura, T. Mitsuhashi, and T. Matsumoto. 1977. Sterols of Liliaceae. *Phytochemistry* 16: 140.
- Itokawa, Y., K. Inoue, S. Sasagawa, and M. Fujiwara. 1973. Effect of S-methylcysteine sulphoxide, S-allylcysteine sulphoxide and related sulfur-containing amino acids on lipid metabolism of experimental hypercholesterolemic rats. *Journal of Nutrition* 103: 88.
- Jain, A. K., R. Vargas, S. Gotzkowsky, and F. G. McMahon. 1993. Can garlic reduce levels of serum lipids? A controlled clinical study. *American Journal of Medicine* 94: 632-5.
- Jones, H. A., and L. K. Mann. 1963. *Onions and their allies; Botany, cultivation and utilization*. London.
- Kaack, K., G. Kjeldsen, and L. Mune. 1993. Changes in quality attributes during growth of leek (*Allium porrum* L.) for industrial processing. *Acta Agriculturae Scandinavica Section B Soil and Plant Science* 43: 172-5.

- Kameoka, H., and S. Hashimoto. 1983. Two sulphur constituents from *Allium schoenoprasum*. *Phytochemistry* 22: 294.
- Karawya, M. S., S. M. Abdel Wahab, M. M. El-Olemy, and N. M. Farrag. 1984. Diphenylamine, an antihyperglycaemic agent from onion and tea. *Journal of Natural Products* 47: 775.
- Kepka, A. K., and M. A. Sypien. 1971. The influence of some factors on the keeping quality of onions. *Acta Horticultura* 20: 65.
- Khodzhaeva, M. A., and E. S. Kondratenko. 1983. *Allium* carbohydrates. III. Characteristics of *Allium* species polysaccharides. *Khimia Prirodnukh Soedinenii* 2: 228.
- Kiesewetter, H., F. Jung, E. M. Jung, et al. 1993. Effect of garlic on platelet aggregation in patients with increased risk of juvenile ischemic attack. *European Journal of Clinical Pharmacology* 45: 333-6.
- Kimura, Y., and K. Yamamoto. 1964. Cytological effects of chemicals on tumours. XXIII. Influence of crude extracts from garlic and some related species on MTK-sarcoma. III. *Gann* 55: 325.
- Kirk, J. H., and M. S. Bulgin. 1979. Effects of feeding cull domestic onions (*Allium cepa*) to sheep. *American Journal of Veterinary Research* 40: 397.
- Knapp, V. J. 1989. Dietary changes and the decline of scurvy and tuberculosis in 19th century Europe. *New York State Journal of Medicine* 89: 621-4.
- Kumar, J., and J. K. Gupta. 1993. Nectar sugar production and honeybee foraging activity in three species of onion (*Allium* species). *Apidologie* 24: 391-6.
- Kumar, K., and R. K. Sahay. 1954. Effect of sulfur fertilization on the pungency of onion. *Current Science* 24: 368.
- Kuroda, S. 1977. Taboo on breeding cloven-hoofed animals at a community in Mujagi prefecture and its influence on dietary habits. *Journal of the Japanese Society for Food and Nutrition* 30: 249.
- Lawson, L. D., D. K. Ransom, and B. G. Hughes. 1992. Inhibition of whole blood platelet aggregation by compounds in garlic clove extracts and commercial garlic products. *Thrombosis Research* 65: 141-56.
- Lewis, N. F., B. Y. K. Rao, A. B. Shah, et al. 1977. Antibacterial activity of volatile components of onion (*Allium cepa*). *Journal of Food Science and Technology (Mysore)* 14: 35.
- Li, H.-L. 1969. The vegetables of ancient China. *Economic Botany* 23: 253.
- Liakopoulou-Kyriakides, M., and Z. Sinakos. 1992. A low molecular weight peptide from *Allium porrum* with inhibitory activity on platelet aggregation in vitro. *Biochemistry International* 28: 373-8.
- Loeper, M., and M. Debray. 1921. Antihypertensive action of garlic extract. *Bulletin of the Society of Medicine* 37: 1032.
- Lohmer, C. 1988. Certain aspects of the nutrition of monks in the Middle Ages with the monastic teaching of Peter Damian as example. *Aktuelle Ernährungsmedizin* 13: 179-82.
- Lucas, R. 1966. *Nature's medicines - the folklore, romance and value of herbal remedies*. New York.
- Lybarger, J. A., J. S. Gallagher, D. W. Pulver, et al. 1982. Occupational asthma induced by inhalation and ingestion of garlic. *Journal of Allergy and Clinical Immunology* 69: 448.
- Mader, F. H. 1990. Treatment of hyperlipidaemia with garlic-powder tablets. Evidence from the German Association of General Practitioners' multicentric placebo-controlled double-blind study. *Arzneimittel-Forschung* 40: 1111-16.
- Mahajan, V. M. 1981. Antimycotic activity of different chemicals, chaksine iodide, and garlic. *Mykosen* 26: 94.
- Makheja, A. N., J. Y. Vanderhoek, and J. M. Bailey. 1979. Properties of an inhibitor of platelet aggregation and thromboxane synthesis isolated from onion and garlic. *Thrombosis and Haemostasis* 42: 74.
- Mathur, P. B. 1963. Extension of storage life of garlic bulbs by gamma-irradiation. *International Journal of Applied Radiation and Isotopes* 14: 625.
- Mauron, J. 1986. Food, mood and health: the medieval outlook. *International Journal for Vitamin and Nutrition Research* 29S: 9-26.
- Mazelis, M., and L. Crews. 1968. Purification of the alliin lyase of garlic, *Allium sativum* L. *Biochemical Journal* 108: 725.
- Mazza, G. 1980. Relative volatilities of some onion flavour components. *Journal of Food Technology* 15: 35.
- Meng, Y., D. Lu, N. Guo, et al. 1993. Studies on the anti-HCMV effect of garlic components. *Virologica Sinica* 8: 147-50.
- Mennella, J. A., and G. K. Beauchamp. 1991. Maternal diet alters the sensory qualities of human milk and the nursing's behavior. *Pediatrics* 88: 737-44.
1993. The effects of repeated exposure to garlic-flavored milk on the nursing's behavior. *Pediatric Research* 34: 805-8.
- Meyer, R. G. 1980. Carbonized feed plants of Pompeii, Herculaneum and the villa at Torre Annunziata. *Economic Botany* 34: 401.
- Miller, B. S., Y. Pomeranz, H. H. Converse, and H. R. Brandenburg. 1977. Removing garlic contamination from harvested wheat. *U.S. Department of Agriculture Product Research Report* 173: 1.
- Mitchell, J. C. 1980. Contact sensitivity to garlic (*Allium*). *Contact Dermatitis* 6: 356.
- Moore, G. S., and R. D. Atkins. 1977. The fungicidal and fungistatic effects of an aqueous garlic extract on medically important yeast-like fungi. *Mycologia* 69: 341.
- Morse, D. L., L. K. Pickard, J. J. Guzewich, et al. 1990. Garlic-in-oil associated botulism: Episode leads to product modification. *American Journal of Public Health* 80: 1372-3.
- Munday, R., and E. Manns. 1994. Comparative toxicity of prop(en)yl disulfides derived from Alliaceae: Possible involvement of 1-propenyl disulfides in onion-induced hemolytic anemia. *Journal of Agricultural and Food Chemistry* 42: 959-62.
- Mutsch-Eckner, M., C. A. J. Erdelmeier, O. Sticher, and H. D. Reuter. 1993. A novel amino acid glycoside and three amino acids from *Allium sativum*. *Journal of Natural Products (Lloydia)* 56: 864-9.
- Naito, S., N. Yamaguchi, and Y. Yokoo. 1981a. Studies on natural antioxidant. II. Antioxidative activities of vegetables of the *Allium* species. *Journal of the Japanese Society for Food Science and Technology* 28: 291.
- 1981b. Studies on natural antioxidant. III. Fractionation of antioxidant extracted from garlic. *Journal of the Japanese Society for Food Science and Technology* 28: 465.
- Nasseh, M. O. 1983. Wirkung von Rotextrakten aus *Allium sativum* L. auf Getreideblattläuse Sitobion avenae F. und Rhopalosiphum padi L. sowie die grüne Pfirsichblattlaus Myzus persicae Sulz. *Angewandte Entomologica* 95: 228.
- National Research Council. 1989. *Recommended dietary allowances*. Tenth revised edition. Washington, D.C.
- Nilsson, T. 1979. Yield, storage ability, quality, and chemical

- composition of carrot, cabbage and leek at conventional and organic fertilizing. *Acta Horticultura* 93: 209.
1980. The influence of the time of harvest on the chemical composition of onions. *Swedish Journal of Agricultural Research* 10: 77.
- Nordestgaard, A. 1983. Growing chives for seed production. *Meddelelse, Statens Planteavltsforsog* 85: 3.
- Nutrition Coordinating Center. 1994. *Nutrition Data System Version 2.6/8A/23*. St. Paul, Minn.
- Odebiyi, A. I. 1989. Food taboos in maternal and child health: The views of traditional healers in Ile-Ife, Nigeria. *Social Science and Medicine* 28: 985-96.
- Oka, Y., S. Kiriyaama, and A. Yoshida. 1974. Sterol composition of spices and cholesterol in vegetable food stuffs. *Journal of the Japanese Society for Food and Nutrition* 27: 347.
- Omar, F. A., and A. E. Arafa. 1979. Chemical composition of garlic bulbs during storage as affected by MH as a pre-harvest foliar spray. *Agricultural Research and Development* 57: 203.
- Omidiji, O. 1993. Flavonol glycosides in the wet scale of the deep purple onion (*Allium cepa* L. cv. Red Creole). *Discovery and Innovation* 5: 139-41.
- Omkumar, R. V., S. M. Kadam, A. Banerji, and T. Ramasarma. 1993. On the involvement of intramolecular protein disulfide in the irreversible inactivation of 3-hydroxy-3-methylglutaryl-CoA reductase by diallyl disulfide. *Biochimica et Biophysica Acta* 1164: 108-12.
- Perchellet, J. P., E. M. Perchellet, N. L. Abney, et al. 1986. Effects of garlic and onion oils on glutathione peroxidase activity, the ratio of reduced/oxidized glutathione and ornithine decarboxylase induction in isolated mouse epidermal cells treated with tumor promoters. *Cancer Biochemistry Biophysics* 8: 299-312.
- Perez, H. A., M. de la Rosa, and R. Apitz. 1994. In vivo activity of ajoene against rodent malaria. *Antimicrobial Agents and Chemotherapy* 38: 337-9.
- Perezgrovas Garza, R. 1990. El uso de la herbolaria como alternativa terapeutica en ovinocultura (The use of medicinal plants as an alternative medicine in sheep farming). *Memoria III Congreso Nacional de Produccion Ovina, Tlaxcala, 25 a 28 de abril 1990*. Universidad Chiapas, Mexico, 242-6.
- Perkin, A. G., and J. J. Hummel. 1896. Occurrence of quercetin in the outer skins of the bulb of the onion (*Allium cepa*). *Journal of the Chemical Society* 69: 1295.
- Peters, E. J., and R. A. Mckelvey. 1982. Herbicides and dates of application for control and eradication of wild garlic (*Allium vineale*). *Weed Science* 30: 557.
- Petkov, V. 1986. Bulgarian traditional medicine: A source of ideas for phytopharmacological investigations. *Journal of Ethnopharmacology* 15: 121-32.
- Piotrowski, G. 1948. L'ail en therapeutique. *Praxis* 48: 8.
- Platenius, H. 1941. Factors affecting onion pungency. *Journal of Agricultural Research* 62: 371.
- Pordesimo, A. N., and L. L. Ilag. 1976. Toxicity of garlic juice to plant pathogenic organisms. *Philippino Journal of Biology* 5: 251.
- Poulsen, K. P., and P. Nielsen. 1979. Freeze drying of chives and parsley - optimization attempts. *Bulletin de Institut International du Froid* 59: 1118.
- Pratt, D. E., and B. M. Watts. 1964. The antioxidant activity of vegetable extracts. I. Flavone aglycones. *Journal of Food Science* 29: 27.
- Priyadarshini, E., and P. G. Tulpule. 1976. Aflatoxin production on irradiated foods. *Food and Cosmetics Toxicology* 14: 293.
- Pruthi, J. S. 1980. Spices and condiments. Chemistry, microbiology, and technology. *Advances in Food Research Supplement* 4: 198.
- Pruthi, J. S., L. J. Singh, and G. Lal. 1959. Thermal stability of alliinase and enzymatic regeneration of flavour in odourless garlic powder. *Current Science* 28: 403.
- Pruthi, J. S., L. J. Singh, S. D. V. Ramu, and G. Lal. 1959. Pilot plant studies on the manufacture of garlic powder. *Food Science* 8: 448-53.
- Pushpendran, C. K., T. P. A. Devasagayam, and J. Eapen. 1982. Age related hyperglycaemic effect of diallyl disulphide in rats. *Indian Journal of Experimental Biology* 20: 428.
- Qureshi, A. A., Z. Z. Din, N. Abuirmeleh, et al. 1983. Suppression of avian hepatic lipid metabolism by solvent extracts of garlic. Impact on serum lipids. *Journal of Nutrition* 113: 1746.
- Raj, K. P. S., Y. K. Agrawal, and M. R. Patel. 1980. Analysis of garlic for its metal contents. *Journal of the Indian Chemical Society* 57: 1121.
- Ramanujam, K. 1962. Garlic in the treatment of acute leprosy neuritis. *Leprosy in India* 34: 174.
- Rao, M. 1985. Food beliefs of rural women during the reproductive years in Dharwad, India. *Ecology of Food and Nutrition* 6: 93-103.
- Reimers, F., S. E. Smolka, S. Werres, et al. 1993. Effect of ajoene, a compound derived from *Allium sativum*, on phytopathogenic and epiphytic micro-organisms. *Zeitschrift für Pflanzenkrankheiten und Pflanzenschutz* 100: 622-33.
- Reznik, P. A., and Y. G. Imbs. 1965. Ixodid ticks and phytocides. *Zoologicheskii Zhurnal* 44: 1861.
- Rick, R. C. 1978. The tomato. *Scientific American* 239: 66.
- Rickard, P. C., and R. Wickens. 1977. The effect of time of harvesting of spring sown dry bulb onions on their yield, keeping ability and skin quality. *Experimental Horticulture* 29: 45.
- Robinson, J. E., K. M. Browne, and W. G. Burton. 1975. Storage characteristics of some vegetables and soft fruits. *Annals of Applied Biology* 81: 399.
- Rosin, S., H. Tuorila, and A. Utela. 1992. Garlic: A sensory pleasure or a social nuisance? *Appetite* 19: 133-43.
- Sainani, G. S., D. B. Desai, N. H. Gorhe, et al. 1979. Effect of dietary garlic and onion on serum lipid profile in Jain community. *Indian Journal of Medical Research* 69: 776.
- Sainani, G. S., D. B. Desai, S. M. Natu, et al. 1979. Dietary garlic, onion and some coagulation parameters in Jain community. *Journal of the Association of Physicians of India* 27: 707.
- San-Blas, G., L. Marino, F. San-Blas, and R. Apitz-Castro. 1993. Effect of ajoene on dimorphism of *Paracoccidioides brasiliensis*. *Journal of Medical and Veterinary Mycology* 31: 133-41.
- Sanchez-Mirt, A., F. Gil, and R. Apitz-Castro. 1993. In vitro inhibitory effect and ultrastructural alterations caused by ajoene on the growth of dematiaceous fungi: *Cladosporium carrionii* and *Fonsecaea pedrosoi*. *Revista Iberoamericana de Micologia* 10: 74-8.
- Sato, A., M. Terao, and M. Ishibashi. 1993. Antibacterial effects of garlic extract on *Vibrio parahaemolyticus* in fish meat. *Journal of the Food Hygienic Society of Japan* 34: 63-7.
- Schreyen, L., P. Dirinck, F. Van Wassenhove, and G. Schamp. 1976a. Analysis of leek volatiles by headspace condensation. *Journal of Agricultural and Food Chemistry* 24: 1147.

- 1976b. Volatile flavor components of leek. *Journal of Agricultural and Food Chemistry* 24: 336.
- Schwimmer, S. 1968. Enzymatic conversion of trans(+)-S-(1-propenyl)-L-cysteine-S-oxide to the bitter and odor-bearing components of onion. *Phytochemistry* 7: 401.
- Sendl, A., M. Schliack, R. Loser, et al. 1992. Inhibition of cholesterol synthesis in vitro by extracts and isolated compounds prepared from garlic and wild garlic. *Atherosclerosis* 94: 79-85.
- Seuri, M., A. Taivanen, P. Ruoppi, and H. Tukiainen. 1993. Three cases of occupational asthma and rhinitis caused by garlic. *Clinical and Experimental Allergy* 23: 1011-14.
- Sharma, A., G. M. Tewari, C. Bandyopadhyay, and S. R. Padwal-Desai. 1979. Inhibition of aflatoxin-producing fungi by onion extracts. *Journal of Food Science* 44: 1545.
- Sharma, K. K., R. K. Gupta, S. Gupta, and K. C. Samuel. 1977. Antihyperglycaemic effect of onion: Effect of fasting blood sugar and induced hyperglycemia in man. *Indian Journal of Medical Research* 65: 422.
- Siddeswaran, K., and C. Ramaswami. 1987. Inter-cropping and border-cropping of compatible crops in finger millet (*Eleusine coracana* Gaertn.) under garden land conditions. *Journal of Agronomy and Crop Science* 158: 246-9.
- Silagy, C., and A. Neil. 1994. Garlic as a lipid lowering agent - a meta-analysis. *Journal of the Royal College of Physicians of London* 28: 39-45.
- Simmonds, N. W. 1976. *Evolution of crop plants*. London.
- Singh, K. K. 1994. Development of a small capacity dryer for vegetables. *Journal of Food Engineering* 21: 19-30.
- Singh, L. J., J. S. Pruthi, A. N. Sankaran, et al. 1959. Effect of type of packaging and storage temperature on flavor and colour of garlic powder. *Food Science* 8: 457-60.
- Singh, L. J., J. S. Pruthi, V. Sreenivasamurthy, et al. 1959. Effect of type of packaging and storage temperature on ally sulphide, total sulfur, antibacterial activity and volatile reducing substances in garlic powder. *Food Science* 8: 453-6.
- Singh, R. V. 1991. Effect of intercrops on performance and production economics of tomato (*Lycopersicon esculentum*). *Indian Journal of Agricultural Sciences* 61: 247-50.
- Smith-Jones, S. 1978. Herbs: Next to his shoes, the runner's best friend may be in the kitchen. *Runner's World* 13: 126-7.
- Soffar, S. A., and G. M. Mokhtar. 1991. Evaluation of the antiparasitic effect of aqueous garlic (*Allium sativum*) extract in hymenolepiasis nana and giardiasis. *Journal of the Egyptian Society of Parasitology* 21: 497-502.
- Sokolov, R. 1975. A plant of ill repute. *Natural History* 84: 70.
- Sreenivasamurthy, V., K. R. Sreekantiah, A. P. Jayaraj, et al. 1962. A preliminary report on the treatment of acute lepromatous neuritis with garlic. *Leprosy in India* 34: 171-3.
- Srinivasan, V. 1969. A new antihypertensive agent. *Lancet* 2: 800.
- St. Louis, M. E., S. H. Peck, D. Bowering, et al. 1988. Botulism from chopped garlic: Delayed recognition of a major outbreak. *Annals of Internal Medicine* 108: 363-8.
- Stallknecht, G. F., J. Garrison, A. J. Walz, et al. 1982. The effect of maleic hydrazide salts on quality and bulb tissue residues of stored "Yellow Sweet Spanish" onions. *Horticultural Science* 17: 926.
- Starke, H., and K. Herrmann. 1976a. Flavonols and flavones of vegetables. VI. On the changes of the flavonols of onions. *Zeitschrift für Lebensmittel Untersuchung und Forschung* 161: 137.
- 1976b. Flavonols and flavones of vegetables. VII. Flavonols of leek, chive, and garlic. *Zeitschrift für Lebensmittel Untersuchung und Forschung* 161: 25-30.
- Stoianova-Ivanova, B., A. Tzutzulova, and R. Caputto. 1980. On the hydrocarbon and sterol composition in the scales and fleshy part of *Allium sativum* Linnaeus bulbs. *Rivista Italiana EPPOS* 62: 373.
- Stoll, A., and E. Seebeck. 1948. Allium compounds. I. Alliin, the true mother compound of garlic oil. *Helvetica Chimica Acta* 31: 189.
- Subrahmanyam, V., K. Krishnamurthy, V. Sreenivasamurthy, and M. Swaminathan. 1957. Effect of garlic in the diet on the intestinal microflora of rats. *Journal of Scientific Indian Research* 160: 173.
- Subrahmanyam, V., V. Sreenivasamurthy, K. Krishnamurthy, and M. Swaminathan. 1957. Studies on the antibacterial activity of spices. *Journal of Scientific Indian Research* 160: 240.
- Sun, Y., J. Sun, X. Liu, et al. 1993. Investigation and experimental research on the effects of onion on angiocardopathy. *Acta Nutrimenta Sinica* 14: 409-13.
- Sutabhaha, S., M. Suttajit, and P. Niyomca. 1992. Studies of aflatoxins in Chiang Mai, Thailand. *Kitasato Archives of Experimental Medicine* 65: 45-52.
- Tatarrintsev, A. V., P. V. Vrzhets, D. E. Ershov, et al. 1992. Ajoene blockade of integrin-dependent processes in the HIV-infected cell system. *Vestnik Rossiiskoi Akademii Meditsinskikh Nauk* O 11/12: 6-10.
- Tesi, R., and A. Ricci. 1982. The effect of plant spacing on garlic production. *Annali della Facolta di Scienze Agrarie della Universita dali Studi di Napoli Portici* 16: 6.
- Thakur, D. E., S. K. Misra, and P. C. Choudhuri. 1983. Trial of some of the plant extracts and chemicals for their antifungal activity in calves. *Indian Veterinary Journal* 60: 799.
- Thamizharasi, V., and P. Narasimham. 1993. Effect of heat treatment on the quality of onions during long-term tropical storage. *International Journal of Food Science and Technology* 28: 397-406.
- Thompson, A. K. 1982. *The storage and handling of onions*. Report G160, Tropical Products Institute. London.
- Thompson, A. K., R. H. Booth, and F. J. Proctor. 1972. Onion storage in the tropics. *Tropical Research* 14: 19.
- Tissut, M. 1974. Étude de la localisation et dosage in vivo des flavones de l'oignon. *Compte Rendu Académie Scientifique Paris Series D* 279: 659.
- Tokarska, B., and K. Karwowska. 1983. The role of sulfur compounds in evaluation of flavoring value of some plant raw materials. *Die Nabrung* 27: 443.
- Toma, R. B., and M. L. Curry. 1980. North Dakota Indians' traditional foods. *Journal of the American Dietetic Association* 76: 589-90.
- Traub, H. L. 1968. The subgenera, sections and subsections of *Allium* L. *Plant Life* 24: 147.
- Urbina, J. A., E. Marchan, K. Lazardi, et al. 1993. Inhibition of phosphatidylcholine biosynthesis and cell proliferation in *Trypanosoma cruzi* by ajoene, an antiplatelet compound isolated from garlic. *Biochemical Pharmacology* 45: 2381-7.
- Urquhart, R., and Y. Webb. 1985. Adverse reactions to food in Down syndrome children and the nutritional consequences. *Proceedings of the Nutrition Society of Australia* 10: 117.
- Usher, G. 1974. *A dictionary of plants used by man*. London.

- Vandenberg, L., and C. P. Lentz. 1974. High humidity storage of some vegetables. *Journal of the Institute of Canadian Science and Technology. Alimentation* 7: 260.
- Van Dijk, P. 1993. Survey and characterization of potyviruses and their strains of *Allium* species. *Netherlands Journal of Plant Pathology* 99(2S): 1-48.
- Van Hecke, E. 1977. Contact allergy to onion. *Contact Dermatitis* 3: 167.
- Van Ketal, W. C., and P. de Haan. 1978. Occupational eczema from garlic and onion. *Contact Dermatitis* 4: 53.
- Van Petten, G. R., W. G. Hilliard, and W. T. Oliver. 1966. Effect of feeding irradiated onion to consecutive generations of the rat. *Food and Cosmetics Toxicology* 4: 593.
- Van Petten, G. R., W. T. Oliver, and W. G. Hilliard. 1966. Effect of feeding irradiated onion to the rat for 1 year. *Food and Cosmetics Toxicology* 4: 585-92.
- Virtanen, A. L. 1965. Studies on organic sulphur compounds and other labile substances in plants - a review. *Phytochemistry* 4: 207.
- Walker, J. C., and M. A. Stahman. 1955. Chemical nature of disease resistance in plants. *Annual Review of Plant Physiology* 6: 351.
- Warren, C. P. W. 1970. Some aspects of medicine in the Greek Bronze Age. *Medical History* 14: 364.
- Warshafsky, S., R. S. Kamer, and S. L. Sivak. 1993. Effect of garlic on total serum cholesterol. A meta-analysis. *Annals of Internal Medicine* 119: 599-605.
- Weber, N. D., D. O. Andersen, J. A. North, et al. 1992. In vitro virucidal effects of *Allium sativum* (garlic) extract and compounds. *Planta Medica* 58: 417-23.
- Wertheim, T. 1845. Investigations of garlic oil. *Annalen der Chemie* 51: 289.
- Whitaker, J. R. 1976. Development of flavor, odor, and pungency in onion and garlic. *Advances in Food Research* 22: 73.
- Wilkens, W. F. 1964. Isolation and identification of the lachrymogenic compound of onion. *Cornell University, Agricultural Experiment Station Memoir*, 385. New York.
- Wills, E. D. 1956. Enzyme inhibition by allicin, the active principle of garlic. *Biochemical Journal* 63: 514.
- Witte, J. S., M. P. Longnecker, C. L. Bird, et al. 1996. Relation of vegetable, fruit, and grain consumption to colorectal adenomatous polyps. *American Journal of Epidemiology* 144: 1015.
- Wright, P. J. 1993. Effects of nitrogen fertilizer, plant maturity at lifting, and water during field-curing on the incidence of bacterial soft rot of onion in store. *New Zealand Journal of Crop and Horticultural Science* 21: 377-81.
- Wright, P. J., C. N. Hale, and R. A. Fullerton. 1993. Effect of husbandry practices and water applications during field curing on the incidence of bacterial soft rot of onions in store. *New Zealand Journal of Crop and Horticultural Science* 21: 161-4.
- Wu, J. L., C. C. Chou, M. H. Chen, and C. M. Wu. 1982. Volatile flavor compounds from shallots. *Journal of Food Science* 47: 606.
- Yamato, O., T. Yoshihara, A. Ichihara, and Y. Maede. 1994. Novel Heinz body hemolysis factors in onion (*Allium cepa*). *Bioscience Biotechnology and Biochemistry* 58: 221-2.
- Yang, G., B. Ji, and Y. Gao. 1993. Diet and nutrients as risk factors of colon cancer: A population-based case control study in Shanghai. *Acta Nutrimenta Sinica* 14: 373-9.
- Yeh, Y. Y., and S. M. Yeh. 1994. Garlic reduces plasma lipids

by inhibiting hepatic cholesterol and triacylglycerol synthesis. *Lipids* 29: 189-93.

Yoshida, S., S. Kasuga, N. Hayashi, et al. 1987. Antifungal activity of ajoene derived from garlic. *Applied and Environmental Microbiology* 53: 615-17.

Yoshikawa, K., K. Hadame, K. Saitoh, and T. Hijikata. 1979. Patch tests with common vegetables in hand dermatitis patients. *Contact Dermatitis* 5: 274.

You, W. C., W. J. Blot, Y. S. Chang, et al. 1989. Allium vegetables and reduced risk of stomach cancer. *Journal of the National Cancer Institute* 81: 162-4.

II.C.3 Beans, Peas, and Lentils

The Names

On Sunday, November 4, 1492, three weeks after his first landing in the New World, Christopher Columbus saw lands planted with "*faxones* and *fabas* very diverse and different from ours [those of Spain] and two days afterward, following the north coast of Cuba," he again found "land well cultivated with these *fexoes* and *habas* much unlike ours" (Hedrick 1931: 3).

In a transcription (Dunn and Kelley 1989: 132) from Columbus's diary, the Spanish phrase *faxones y favas* has been translated as "beans and kidney beans" (Morrison and Jane-Vigneras, cited by Dunn and Kelley 1989: 133). But considering what Columbus might have seen in the markets or kitchens of the fifteenth-century Iberian-Mediterranean world, *faxone* probably refers to the African-Asian cowpea (*Vigna unguiculata*), and *fava* surely means the fava (= *faba*), or broad bean (*Vicia faba*), long known in Europe and the Mediterranean-Asian world. Columbus's brief record presaged the long confusion and debate over the names and origins of some important food grain legumes. Had herbalists and botanical authors of the succeeding three centuries taken account of Columbus's recognition that these New World legumes were different from those of Europe, some of the confusion might have been avoided.

The beans, peas, and lentils (pulses, or food grain legumes) discussed in this chapter are legumes, treated in technical botanical literature as members of the very large family Fabaceae (= Leguminosae), subfamily Papilionoideae (having butterflylike flowers), although some taxonomists accord this group family status (Papilionaceae). The names of the species, however, are not changed by the differing positions taken by plant taxonomists on family nomenclature. The flowers of papilionaceous legumes have five petals consisting of one broad standard, two lateral wings, and two keel petals that clasp the 10 stamens and single ovary (which becomes the pod).

Lentil



The Latin and most frequently used English common names of the species of beans, peas, and lentils discussed in this chapter are enumerated in Table II.C.3.1, along with other species having similar names.

The European herbalists of the sixteenth century repeatedly attempted to reconcile their plant world with that of the ancients as represented by the fragmentary remains of the works of Theophrastus and Dioscorides. Their efforts, however, were destined to fail inasmuch as many of their subjects were novel plants newly introduced from the New World and from the reopened contact with Asia and coastal Africa. This nomenclatural dilemma is illustrated by an attempt to reconcile the name of a bean with the appropriate plant. J. P. Tournefort (1656–1708), a notable pre-Linnaean French botanist, sought to clarify the use of *boona* or *baiana* for the broad bean or fava by a well-known sixteenth-century herbalist:

Dodonaeus [said Tournefort] called this kind of pulse “boona” in Latin, who [Dodonaeus] relying on a Germanism abuses his own language in order to appear learned but our Boona or Bean

seem rather to be derived from the Italian word *Baiana* which Hermolaus says is the word used by those that sell new BEANS all over the state of Milan and along the Appenine mountains. . . . Garden beans are common and universal in Europe and are a great supply in dearth of Provisions in the spring and whole summer season. . . . The ancients and Dodonaeus believed that beans are windy and the greener the more so. (Tournefort 1730: 386)

Tournefort then disagreed on the suitability of beans in the diet and said that he would leave the hard dry beans to “the laboring men who can better digest them, but [even] those of delicate constitution and sedentary life digest green beans well enough if they eat them with butter and pepper [especially if they will] be at the pain to take off their skins” (Tournefort 1730: 386). Inasmuch as American *Phaseolus* beans had entered Europe by his lifetime, one could wonder whether Tournefort meant *Phaseolus* or *Vicia* “beans.” However, his remark concerning the removal of “skins [seed coats]” should end any doubt.

Table II.C.3.1. Beans, peas, and lentils

Latin names	English common names
<i>Lens culinaris</i> Meddick.	lentil
<i>Phaseolus acutifolius</i> (A. Gray)	tepany
<i>p. coccineus</i> L. (syn. <i>P. multiflorus</i> Willd.)	scarlet runner, Windsor bean
<i>P. lunatus</i> L.	lima, butter, sieva bean
<i>P. polyanthus</i> (Greenm.)	polyanthus bean
<i>P. vulgaris</i> L.	common or kidney bean, haricot
<i>Pisum sativum</i> L.	pea, garden pea
<i>Vicia faba</i> L.	broad bean, horse bean, fava, haba
Other taxa mentioned in text	
<i>Cicer arietinum</i> L.	chickpea, bengal gram
<i>Lablab purpureus</i> (L.) Sweet (syn. <i>Dolichos lablab</i> L.)	hyacinth bean
<i>Lathyrus sativus</i> L.	chickling vetch, Khesari dhal
<i>Lens culinaris</i> ssp. <i>macrosperma</i> (Baumb)	large-seeded lentil
Barulina	
<i>L. culinaris</i> ssp. <i>microsperma</i>	small-seeded lentil
Barulina	
<i>L. ervoides</i> (Brign.) Grande	wild lentil erse, French lentil cowpea, black-eyed pea
<i>L. nigricans</i> (M. Bieb.) Godr.,	
<i>L. orientalis</i> (Bioss.) M. Popov	
<i>Vicia ervilia</i> Willd.	
<i>V. sinensis</i> (L.) Savi ex Hassk.	
<i>Vigna unguiculata</i> (L.) Walp. ssp. <i>unguiculata</i> (syn. <i>V. Sesquipedalis</i> [L.] Frubw.)	

Sources: Adapted from Aykroyd and Doughty (1964), pp. 15–18, Purselove (1968), Smartt (1990), Ladizinsky (1993), and other sources.

"Skins" refers to the seed coats. Where the broad bean and American *Phaseolus* beans are concerned, only the broad bean has its skin or testa customarily removed – probably to eliminate substances that are toxic for individuals having an inherited enzyme (glucose-6-phosphate dehydrogenase) deficiency. Even in a contemporary and remote Quechua-speaking community, located in the southern Andes in the vicinity of Cuzco at an altitude of 3,000 to 5,000 meters, which is unfavorable for *Phaseolus* cultivation, cooked *Vicia faba* seeds are always peeled before eating (Franquemont et al. 1990: 83). Because this enzyme-deficiency sensitivity to fava bean components evolved in human populations native to certain malarial regions of Eurasia and Africa (Strickberger 1985: 738), the custom in the Andes was probably introduced along with the fava bean.

Tournefort further assumed the responsibility for ending the quandary of post-Columbian botanists concerning the identity of the fava. He recognized that there was

... much controversy among the botanists as to whether our bean be the bean of the ancients ... that of the ancients was small and round [according to] Theophrastus, Dioscorides and others. But it is strange that a pulse so common should have come close to disuse and been replaced without anybody's knowing anything of this matter. (Tournefort 1730: 386)

The reason for the difference (and confusion), he went on, could be that "their faba was not arrived at the bigness that our Garden Bean now is." But however intriguing the evolutionary explanation, the writers of the classical period, whom he cites, may have been referring to the African-Asian cowpea, *Vigna unguiculata*, rather than one of the small-seeded races of the fava bean.

Contemporary linguistic sources (*Webster's Third New International Dictionary* 1971) derive the familiar English word "bean" from a root common to Old English, Old High German, and Old Norse that apparently once referred to the fava or faba bean, a staple of the Romans. Over the centuries, however, the word has grown to encompass seeds of other plants, including a multitude of legumes besides the fava, most of which belong to other genera – a terminological tangle that botany attempts to avoid through the use of scientific names (Table II.C.3.1).

The distinct identities of these two groups of food crops, favas and *Phaseolus* beans, were being established as seventeenth- and eighteenth-century botany ceased the attempt to reconcile the known species of that period with the fragmentary records of the classical authors. These advances were only the beginning of the solutions to the geographic, temporal, and cultural problems surrounding the origins of these foods.

The Search for Geographic Origins

The origins of most domesticated plants, meaning the time span, the wild relatives, and the conditions (both natural and human-influenced) under which divergences from wild ancestral stock took place, are remote in time and are matters of continued botanical and genetic inquiry.

For Europe and the Mediterranean Basin, sixteenth-century European herbalists turned for information to tradition, the observations of travelers, and the surviving books of classical authors. Linnaeus assimilated the writings of the herbalists and added his contemporary experience with eighteenth-century introductions and collections. Alphonse de Candolle (1964: v) in 1855, and especially in the 1886 edition of his *Origin of Cultivated Plants*, brought to the attention of botanists the utility of archaeological findings for supplementing plant morphology and taxonomy in adducing evidence for the geography of domestication.

In the twentieth century, N. I. Vavilov, following Candolle's pioneering work in geographical botany, embarked on his global, decades-long projects on the origin and genetic variation of cultivated plants. In 1926, Vavilov (1992: 22-7) organized a comparative chart of the morphology of cultivated species of the papilionaceous legumes and presented the rationale for botanical-genetic determination of the centers of origin of these and other crop plants. His geographic centers of origin for crop plants – which have been highly influential and much discussed in the literature of crop plant geography – included eight major and several minor centers (Vavilov 1992: 324-53). The lentil, the pea, and the broad bean were all traced to the Inner-Asiatic Center: northwestern India, Afghanistan, Tadzhikistan, Uzbekistan, and western China. The lentil and pea were also assigned by Vavilov to the Asia Minor Center (the Middle East, Iran, and Turkmenistan). The American common bean and lima bean were both assigned primary centers in the South Mexican-Central American Center, and the lesser-known species, scarlet runner beans and teparies, were also located there. The origin of the little-known polyanthus bean (*Phaseolus polyanthus*) has been documented in the same area (Schmit and Debouck 1991).

New World Beans: *Phaseolus*

The Pathway of Domestication in *Phaseolus*

The species of domesticated *Phaseolus* beans have shown parallel changes in structure and physiology and share some of these characteristics with the Old World grain legumes. J. Smartt (1990: 111) summarized his own findings and those of others on the nature of evolutionary changes in the *Phaseolus* cultigens. These are gigantism (increased size of seed and other plant parts); suppression of seed dispersal mechanisms (decreased tendency of pods to twist

and discharge seeds); changed growth form (especially the loss of rampant vining); loss of seed dormancy; and other physiological and biochemical changes. The genetic bases for seed size, dispersal mechanisms, and growth form are partly understood, and some are even observable in archaeological specimens, in which wild *Phaseolus* beans can be readily distinguished from domesticates by both seed size and nondehiscent pod structure.

The common bean. It is clear in the writings of sixteenth-century herbalists, and later in the works of Linnaeus and Candolle, that the original home of the common bean was unknown to them. It was archaeological excavation on the arid coast of Peru in the last quarter of the nineteenth century that convinced Candolle (1964: 341–2) that *Phaseolus vulgaris* and *Phaseolus lunatus* had been cultivated in the Americas since pre-Columbian times.

At the time of contact with Europeans, varieties of the common bean were grown by Native Americans as far south as Chile and Argentina, and as far north as the valleys of the St. Lawrence and upper Missouri rivers. Edward Lewis Sturtevant (Hedrick 1972: 424) has noted that beans were observed to be in cultivation among Florida Indians by at least three explorers from 1528 to 1562, including Hernando de Soto (in 1539), who said that “the granaries were full of maes and small beans.” The Natchez on the lower Mississippi grew beans as a “subsidiary crop” (Spencer et al. 1965: 410), and there is a 1917 description of traditional Hidatsa–Mandan cultivation of beans in hills between the rows of maize or occasionally “planted separately” (Spencer et al. 1965: 343). This observation suggests the planting of erect or semierect beans. Such beans are intermediate between strong climbers and truly dwarf nonclimbing beans. In California, outside of the lower Colorado River drainage where the Mohave grew tepary beans (as did other Yumans, along with maize, cucurbits, and other crops), bean agriculture began with the introduction of Mexican and European crops when the earliest Spanish missions were established. G. W. Hendry (1934) found a bit of seed coat of the common bean cultivar ‘Red Mexican’, or ‘Pink’, in a Spanish adobe brick dated 1791 at Soledad.

R. L. Beals (1932) mapped pre-1750 bean cultivation in northern Mexico and the adjacent southwestern and Gulf Coast United States using historical documents and reports by Spanish explorers. Bean distribution coincided with maize cultivation, extending from the Colorado River east to include the Rio Grande Pueblos, Zuni, and Hopi. The area of the eastern Apache in the Pecos River drainage, and of the Comanche, was nonagricultural. Beans were grown from eastern Texas, beginning with the Waco and the Kichai, eastward to the Atlantic.

Southwestern bean horticulture derives from Mexico. P. A. Gepts’s (1988: 230) mapping of bean disper-

sal routes by means of the beans’ protein structure corroborates this generally accepted view. Where water is the limiting factor in growth and production, as on the Hopi mesas, varieties having the dwarf or bush habit are planted. Where surface water is available or can be supplied, vining beans are planted with corn or, if planted separately, are provided with poles to climb upon. Except for the pinto or *garapata* group, the common beans collected during the years 1910 to 1912 by G. F. Freeman (1912) among the Papago and Pima do not appear in the archaeology of the southwestern United States. Instead, these beans represent introductions from the Mexican Central Highlands and may well have arrived during the Spanish colonial period.

According to E. F. Castetter and W. H. Bell (1942), the Papago planted teparies in July and harvested in October; the Pima planted twice, when the mesquite leafed out (late March to mid-April) and after the Saguaro harvest (July). The first planting was harvested in June, the second in October. The harvest of teparies and common beans was women’s work. The plants were pulled, dried, and threshed on a clean, hard-packed soil floor in the open, in the field, or near the house. Different varieties were planted, harvested, and threshed separately. After threshing they were sealed in baskets or pots for protection from pests. Castetter and Bell (1942) and Freeman (1912) reported only bush beans in which the “vines” (sprawling plants) are grown separately from corn.

Vavilov (1992) and C. B. Heiser (1965) speculated on multiple American origins for the common bean, and substantial evidence has been adduced to show that common beans were domesticated independently in at least two distinct areas: Mesoamerica and Andean America (Kaplan 1981; Delgado Salinas 1988; Gepts 1988).

Lima beans. The lima and sieva beans were recognized by Linnaeus to belong to the same species, which he called *P. lunatus* to describe the “lunar” shape of the seeds of some varieties. The small-seeded or sieva types are natives of Mexico and Central America. These are distinct from the large-seeded South American lima beans, but can be crossed with them, and for this reason are considered to be of the same species. The sievas appear in the archaeological records of Mexico about 1,200 years ago, but do not occur in the known records of Andean archaeology. The large limas of South America, conversely, do not appear in the Mesoamerican or North American record. Seeds of the South American group have been found in Guitarrero Cave, the same Andean cave as the earliest common beans of South America, and have been 14-Carbon dated at the National Science Foundation, University of Arizona Accelerator Facility, to approximately 3,500 years ago (Kaplan 1995). The archaeological evidence of geographic separation coincides with contemporary observations of the dis-

tribution of both the wild and cultivated types. It seems clear that the two groups were domesticated independently. Both vining and bush forms are known, but the vining forms predominate in indigenous horticulture.

On the desert north coast of Peru, remains of the large-seeded lima group, dating to about 5,600 years ago (the preceramic period) (Kaplan 1995), were well preserved by the arid conditions and were so abundant that they must have constituted a major part of the diet. Ancient Peruvians even included depictions of beans in their imaginative painted ceramics and woven textiles. Painted pottery from the Mochica culture (A.D. 100–800) depicts running messengers, each carrying a small bag decorated with pictures of lima beans (*pallares*). Rafael Larco Hoyle (1943) concluded that the lima beans, painted with parallel lines, broken lines, points, and circles, were ideograms. Some of the beans that Larco Hoyle believed to be painted were certainly renderings of naturally variegated seed coats. Other depictions show stick-limbed bean warriors rushing to the attack. Textiles from Paracas, an earlier coastal site, are rich in bean depictions.

Scarlet runner beans. The cultivated scarlet runner bean (*Phaseolus coccineus* L., for the scarlet flowers) is not known in the archaeological record north of Durango, Mexico, where it was grown about 1,300 years ago (Brooks et al. 1962). It has undoubtedly been cultivated for a much longer period in the cool central highlands of Mexico and Guatemala, but in that region, archaeological specimens securely dated and identified as older than a few hundred years are wanting. Runner beans, both purple-seeded and white-seeded, have been collected among the Hopi in historic times, especially by Alfred Whiting during the 1930s.

Tepary beans. The tepary is one of the two cultivated species not to have been named by Linnaeus. The wild type was called *Phaseolus acutifolius* by the nineteenth-century Harvard botanist Asa Gray. However, it was not until the early years of the twentieth century that the cultivated types were recognized by the Arizona botanist, Freeman, to belong to Gray's species rather than simply being varieties of the common bean.

Teparies, now little-known as commercial beans, were cultivated in central Mexico 2,300 years ago (Kaplan 1994), and in Arizona, teparies were being grown 1,000 to 1,200 years ago (Kaplan 1967). Despite their antiquity and ancient distribution, teparies have been absent from village agriculture in historic times, except in the Sonoran Desert biome of northwestern Mexico, Arizona, and New Mexico, and in the area around Tapachula in the Mexican state of Chiapas and in adjacent Guatemala. They are grown and eaten by the Pima, Papago, and peoples of the

lower Colorado River and by some "Anglo" enthusiasts for dryland-adapted crops. Because of their drought tolerance, they have been tested in many arid regions of the world.

Polyanthus beans. A fifth cultivated bean species, *P. polyanthus*, has long been known to be distinct from the other, better-known species, but only recently have its identity and distribution been documented (Schmit and Debouck 1991). To the best of my knowledge this bean of high elevations in Mexico and Central America has never entered into Old World cultivation.

The five *Phaseolus* beans are distinct species. They have different botanical names, applied by plant systematists on the basis of their structural differences. They have the same number of chromosomes ($2n = 22$) but do not freely hybridize. However, the domesticates do hybridize with some wild-growing populations that are regarded as their ancestral relatives.

The Antiquity of Phaseolus Beans: New Evidence

Uncovering the botanical and geographic origins of domesticated crops includes the search for their temporal origins. Candolle (1964), as noted previously, brought to the attention of plant scientists the utility of archaeological evidence in the quest for the temporal as well as the geographic origins of crop plants. The presence of *Phaseolus* beans on the arid coast of Peru in pre-Conquest graves of indigenous peoples did not indicate a specific calendar date for the remains, but it was convincing evidence that *Phaseolus* beans were present in the Americas before contact with European cultures. With the development of radiometric dating by the middle of the twentieth century, it became possible to determine the age of archaeological organic materials with significant precision. However, many of the published dates for the earliest crop plant remains, including beans (Kaplan 1981), are now being questioned because the 14-Carbon determinations of age are "contextual dates," meaning they are based on organic materials in the same strata with the bean remains but not on the beans themselves (Kaplan 1994).

Because of the tendency of small objects, like seeds, to move downward in archaeological contexts, some of the dates now in the literature are too early. The development of 14-Carbon dating by Atomic Mass Spectrometry (AMS), which measures very small samples, has allowed the dating of single bean seeds or pods, and in some instances has produced dates that disagree with the contextual dates. For example, a single bean-pod found in Coxcatlan Cave in the Tehuacan valley, in a context 6,000 to 7,000 years old, was AMS-dated to only $2,285 \pm 60$ years ago (Kaplan 1967, 1994). An early date for beans in South America comes from Guitarrero Cave

in the Peruvian Andes, where radiocarbon dates of plant debris are as old as 8,000 years (Kaplan, Lynch, and Smith 1973; Lynch et al. 1985). But the AMS 14-Carbon date for one seed embedded in this debris is $2,430 \pm 60$ years old (Kaplan 1994). Disagreements over the accuracy of AMS dates (unpublished data) versus contextual 14-Carbon dates are being aired, and the debate will continue. However, an AMS date from New England (Bendremer, Kellogg, and Largy 1991) supports a contextual 14-Carbon date, which suggests the entry of common beans into northeastern North America (Ohio) about 1,000 years ago (Kaplan 1970). In the southwestern United States, AMS dates (Wills 1988: 128) of beans from a New Mexico cave agree with contextual dates (Kaplan 1981).

The wild types of all *Phaseolus* bean species are vining types, as the earliest domesticates must also have been; the dwarf-growing types came later. The earliest evidence now available for the presence of the dwarf, or bush, growth habit comes from an accelerator radiocarbon date of 1285 ± 55 years ago for bean-plant remains from an archaeological cave site in the northeastern Mexican state of Tamaulipas. The vining or pole beans of each species were planted with corn so that they could depend on the stalks for support. The sprawling and dwarf types could be grown independent of support.

Tracing Bean Migrations

Molecular evidence. Phaseolin is the principal storage protein of *Phaseolus* bean seeds. Gepts (1988) has used the variation in phaseolin structure to trace the dispersal of contemporary common bean cultivars within the Americas and in the Old World. He has shown that the majority of the present-day cultivars of Western Europe, the Iberian Peninsula, Malawian Africa, and the northeastern United States originated in the Andes. The 'C' phaseolin type is found in highest frequency in the Americas in Chile and in the Iberian Peninsula among those Old World regions sampled.

Gepts has applied the phaseolin-structure method to questions of dispersal within the Americas, such as that of the traditional beans of the northeastern United States. There, a majority of the cultivated bean varieties of historic times are of the 'T' phaseolin type, which is the type that is most frequent in Western Europe and in the Andes south of Colombia. 'T' phaseolin is common elsewhere in South America but not in Mesoamerica. Indeed, in Mesoamerica and the adjacent southwestern United States, 'T' types make up only 8 percent and 2 percent, respectively, of the cultivated bean varieties. The archaeological record of crop plants in the northeastern United States is limited because of poor conditions for preservation (humid soils and no sheltered cave deposits), but those beans that have been found, although carbonized, are recognizable as a south-

western United States type (Kaplan 1970). This common bean type, which was dispersed from northwestern Arizona along with eight-rowed corn, must have been of the 'S' phaseolin type, which Gepts has found characteristic of 98 percent of contemporary southwestern common beans. It seems clear that historic-period northeastern bean cultivars are primarily South American, which could have reached the northeastern United States by way of sailing ships, directly from Peruvian and Chilean ports during the late eighteenth and early nineteenth centuries, or from England and France along with immigrants, or through seed importation.

In the foregoing, we see a dispersal pattern that was probably common in much of the New World, and especially in semiarid places in Mesoamerica and the greater Southwest. In dryland prehistoric sites, organic remains are well preserved in the archaeological record, and we see that prehistoric bean cultivars have often been eliminated, or reduced in frequency, by better-adapted (biologically, culturally, economically), introduced cultivars. Such a pattern suggests that Columbus's introduction of beans to Europe from the Caribbean Islands was soon augmented by later introductions from Andean agriculture bearing the 'C' phaseolin type.

Historical evidence. Success in tracing the dispersion of beans from their regions of origin rests to some extent on historical records. But such records are likely to be strongly biased. One of the richest sources of evidence is the body of data available from seed catalogs, magazines, and newspapers devoted to agriculture and horticulture. Such publications, however, are unevenly representative of the larger dispersion picture. The United States, parts of Latin America, and Western Europe may be better represented in this respect than are some other parts of the world. Specialized libraries and archives have preserved some of this published material in the United States for about 200 years. In the United States, the earliest sources for named varieties of garden plants are leaflets or advertisements for seeds newly arrived from England. In Portsmouth, New Hampshire, over a period from 1750 to 1784, 32 varieties of beans, including common beans, both vining and erect types, scarlet runner beans, possibly small-seeded limas, and fava beans, were listed for sale in the *New Hampshire Gazette* (documents courtesy of Strawberry Banke Museum, Portsmouth, N.H.). Earlier still, but less informative, are lists prepared for the guidance of colonists heading for English North America in the seventeenth century, who were advised to supply themselves with peas and (broad) beans, in addition to other vegetable seeds, garden tools, and weapons. We do not begin to detect evidence for the ingress of *Phaseolus* bean cultivars from England and France to the United States until the early nineteenth century.

The Old World: Broad Beans, Peas, and Lentils

As in much of the Americas, the Mediterranean world's combination of cereal grain and food grain legumes has been the foundation for both agriculture and the diet of settled farming communities.

William J. Darby and colleagues (Darby, Ghaloungui, and Grivetti 1977), in tracing food use in Pharaonic Egypt, found papyrus texts and archaeological remains to contain much evidence of food grain legumes in the daily life of the kingdom. Rameses II spoke of barley and beans in immense quantities; Rameses III offered to the Nile god 11,998 jars of "shelled beans." The term "'bean meal' [medicinal bean meal], so commonly encountered in the medical papyri," could apply to *Vicia faba*, to other legumes, or to the "Egyptian bean" (*Nelumbo nucifera* Gaertner, the sacred lotus), all of which have been found in tombs (Darby et al. 1977: II 683).

Fava beans were avoided by priests and others, but the reasons are not clear. The avoidance of favas by Pythagoras, who was trained by Egyptians, is well known, and it may have been that he shared with them a belief that beans "were produced from the same matter as man" (Darby et al. 1977: II 683). Other ancient writers gave different reasons for the taboo, such as self-denial by priests of a variety of foods, including lentils. And more recently, as already mentioned, favism, a genetically determined sensitivity (red blood cells deficient in glucose-6-phosphate dehydrogenase) to chemical components of fava beans, has suggested still another explanation for the avoidance of this food (Sokolov 1984).

Domestication

Structural change under domestication in fava beans, peas, and lentils is much like that in the *Phaseolus* beans, but there are important differences in what can be determined from archaeological samples. Pods of pulses in Middle Eastern sites are seldom found; hence, the loss of pod dehiscence has not been traced in archaeological materials from that region as it has in the Americas (Kaplan and MacNeish 1960; Kaplan 1986). In an effort to find good characters for distinguishing between wild and domesticated types in Middle East pulses, A. Butler (1989) has studied their seed coat structure but has found that the loss of seed dormancy (the impermeability of seed coats to water, which separates the wild types from the domesticates) cannot readily be detected by examination of the seed coats with scanning electron microscopy, or, at best, there is no single structural change that accounts for the difference between permeability and impermeability. Butler (1989: 402) notes that the surface of testa has been used to distinguish wild from cultivated taxa in Viciae. With the exception of testa thickness in seeds of *Pisum*, no characters recorded in the seed coat of Viciae can be associated directly with the presence or absence of hard seed coats. Evi-

dence from seed anatomy of dormancy, and therefore of wildness, is lacking in Viciae.

The origins and domestication of broad beans, peas, and lentils have been the focus of extensive research by plant geneticists and by archaeologists seeking to understand the foundations of agriculture in the Near East. D. Zohary (1989a: 358-63) has presented both genetic and archaeological evidence (not uncontested) for the simultaneous, or near simultaneous, domestication in the Near East of emmer wheat (*Triticum turgidum* ssp. *dicoccum*), barley (*Hordeum vulgare*), and einkorn wheat (*Triticum monococcum*), "hand in hand with the introduction into cultivation of five companion plants": pea (*Pisum sativum*), lentil (*Lens culinaris*), chickpea (*Cicer arietinum*), bitter vetch (*Vicia ervilia*), and flax (*Linum usitatissimum*). *V. faba* may also have been among these earliest domesticates (Kislev 1985).

Lentil

Lens culinaris (*Lens esculenta*), the cultivated lentil, is a widely grown species in a small genus. Archaeological evidence shows that it or its relatives were gathered by 13,000 to 9,500 years ago in Greece (Hansen 1992) and by 10,000 to 9,500 years ago in the Near East (Zohary and Hopf 1988). Candolle (1964) in 1885 wrote of its presence in the Bronze Age sites (the so-called Swiss Lake dwellings) in Switzerland. Lentil cultivation, by historic times, was important and well established in the Near East, North Africa, France, and Spain (Hedrick 1972), and had been introduced into most of the subtropical and warm temperate regions of the known world (Duke 1981: 111). With a production of about 38 percent of the world's lentils and little export (Singh and Singh 1992), the Indian subcontinent - a region in which food grain legumes are an especially important source of protein in the population's largely vegetarian diet - has made this crop one of its most important dietary pulses. Traditionally, two subspecies are recognized on the basis of seed size: *L. culinaris* ssp. *microsperma*, which are small-seeded, and *L. culinaris* ssp. *macrosperma*, which are large-seeded. The large-seeded form is grown in the Mediterranean Basin, in Africa, and in Asia Minor. The small-seeded form is grown in western and southwestern Asia, especially India (Duke 1981: 110-13).

According to G. Ladizinsky (1989: 377, 380) the genus *Lens* comprises *L. culinaris* and four wild species: *Lens nigricans*, *Lens ervoides*, *Lens odemensis*, and *Lens orientalis*. The same author, however, notes that the genus may be reclassified to reflect the genetic relationships of the species, thus: *L. culinaris* ssp. *odemensis*, ssp. *orientalis*, ssp. *culinaris*; *L. nigricans* ssp. *ervoides*, ssp. *nigricans*. Many populations of ssp. *orientalis* and *odemensis* are sufficiently close to ssp. *culinaris* to be used for breeding purposes. However, the chromosome structure within these

subspecies indicates that the cultivated lentils evolved from the principal cytogenetic stock of ssp. *orientalis*. Because this cytogenetic stock ranges from Turkey to Uzbekistan (Ladizinsky et al. 1984), there is little evidence of where in this vast region domestication might have first occurred.

Lentil domestication. Ladizinsky (1987) has maintained that the gathering of wild lentils, and perhaps other grain legumes, prior to their cultivation could have resulted in reduced or lost seed dormancy, a factor of primary importance in the domestication of legumes. As an operational definition of cultivation, Ladizinsky accepted the proposal of Jack R. Harlan, who emphasized human activities designed to manage the productivity of crops and gathered plants. As part of the gathering or foraging process, such management could lead to changes in genetic structure and morphology of wild plants.

High rates of seed dormancy observed by Ladizinsky in wild *L. orientalis* populations demonstrated that they are ill-adapted to cultivation. Dormancy, a single-gene recessive trait, is absent from most of the lineages of domesticated *L. culinaris*. Nondormancy is determined by a single dominant allele. A loss of seed dormancy in lentil populations resulting from gathering practices, Ladizinsky (1979, 1987, 1993) has argued, is the result of establishing this mutant dominant allele in small populations where most of the lentil seeds produced are removed by human gathering. A patchy distribution of small populations of annual legume plants, low population density, and intensive gathering would be, under the conditions he defines, advantageous for the increase of the seed-dormancy-free trait. He has concluded that this process would have taken place in wild, noncultivated lentil (and probably in wild pea and chickpea) populations, and would have predisposed these populations to domestication. This, he maintained, was an evolutionary pathway different from that of the cereal grains for which dormancy of seeds is not a barrier to domestication.

Ladizinsky's views have been disputed by Zohary (1989b) and by M. A. Blumler (1991), who both criticized Ladizinsky's assumptions and conclusions. Blumler investigated Ladizinsky's mathematical model for the loss of dormancy in lentils and concluded that fixation of alleles for nondormancy, as a result of intensive human gathering, did not depend upon gathering at all. He further concluded that Ladizinsky's model would predict loss of dormancy under any circumstances and therefore was not tenable as a hypothesis for preagricultural domestication. He went on to propose that lentils, and possibly other legumes that were gathered and brought to camps, might inadvertently have been introduced to campground soils, but that the poorly competitive introductions would have had to be encouraged (cultivated) in order to survive. Blumler thus contested Ladizinsky's proposal

that legumes could have been domesticated prior to actual cultivation and agreed with Zohary that the pathways of domestication in cereals and legumes were similar, and that legume cultivation followed upon the beginning of cereal cultivation.

As noted earlier in this chapter, the virtual absence of grain legume pods from archaeological sites in the Near East contrasts with the record of *Phaseolus* in the Americas. This circumstance makes it difficult to judge the merit of Ladizinsky's view that the forceful dissemination of seeds following pod dehiscence – a process resulting in the loss of much of a valuable food resource – can be circumvented by pulling whole plants. This would be one sort of gathering technique that could be considered predomesticative management. Accordingly, Ladizinsky placed a low value on the loss of pod dehiscence, or shattering in legumes, while agreeing with most of those concerned that the loss of shattering in cereal grains is of high value in the process of domestication.

Zohary, in response (1989b), accorded the suppression of the wild type of seed dispersal in grain legumes equal weight with the parallel process in the cereal grains. In so doing, he reinforced the argument that domestication in the Near Eastern cereal grains and grain legumes were similar processes. These provocative models for the transition of legumes, and other economic plants, from the wild to the domesticated state provide the basis for both field and theoretical tests. As more evidence is gathered, these theories will be revised repeatedly.

Because archaeological legume pods are seldom found among the plant remains of early Near Eastern societies, dehiscence versus nondehiscence of pods cannot be determined as it has been from New World bean remains. The appearance of seed nondormancy as a marker of the domesticated legume has not been made on the basis of archaeological materials for reasons noted previously in connection with Butler's study of seed coat structure. The archaeological record, however, does reveal changes in legume seed size in the Near East, such as small-seeded (diameter 2.5 to 3.0 millimeters) forms of lentil in early (seventh millennium B.C.) aceramic farming villages. By 5500 to 5000 B.C., lentils of 4.2 millimeters in diameter were present in the Deh Luran Valley of Iran, as reported by Hans Helbaeck (cited by Zohary and Hopf 1988: 89). Zohary and M. Hopf regarded the contrast in size between lentils of the early farming villages and those of 1,500 to 2,000 years later as a change under domestication.

Lentils spread into Europe with the extension of agriculture from the Near East in the sixth millennium B.C., where records of this crop and other legumes during the Bronze Age are less abundant than they are in the earlier Neolithic and later Iron Ages (Zohary and Hopf 1988: 87–9). Only their difference in size from the wild types (*L. orientalis* or *L. nigricans*) gives evidence for the development of

domesticated lentils, which probably occurred in the Near East (Zohary and Hopf 1988: 91–2). These larger lentils appeared in the Near East as late as the end of the sixth millennium B.C., about 1,500 years after the establishment of wheat and barley cultivation (Zohary and Hopf 1988: 91). Although the archaeological and morphological evidence do not disclose the place of origin of the lentil, it is in the Near East where it occurs early in the large-seeded form and where *L. orientalis* is to be found. The lentil, however, is not mentioned as a food crop by E. J. Simoons (1991) in his comprehensive work on food in China.

Pea

Ever a source of confusion for readers of crop plant history and ethnobotany, common names often serve less to identify than to confound. “Pea” (*P. sativum*), as well as “bean,” illustrates this problem. A name with Latin and Greek origins, “pea,” in English, was formed from “pease,” which took on the meaning of a plural; the “-se” was dropped by A.D. 1600 and the current form of the word was established (Oxford English Dictionary 1971). Pea is used combined with descriptors to form the names of numerous plants. In the context of this discussion of food grain legume origins, the important distinction is that between *P. sativum* and the grass pea or chickling pea (*Lathyrus sativus* L.), and the cowpea, black-eye pea, or black-eye bean *Vigna unguiculata*.

The grass pea is a minor crop in the Mediterranean Basin, North Africa, the Middle East, and India, and has been documented archaeologically in parts of this broad region since Neolithic times (Zohary and Hopf 1988: 109–10). In India, it is a weedy growth in cereal grain fields and is the cheapest pulse available (Purseglove 1968: 278). During times of famine, it has been consumed in significant amounts even though its consumption has caused a paralysis of the limb joints (Aykroyd and Doughty 1964).

The cowpea, an important crop in the tropics and subtropics, especially in Africa (Purseglove 1968: 322) and Southeast Asia (Herklots 1972: 261), appeared in the Mediterranean Basin in classical times (Zohary and Hopf 1988: 84). A plant mentioned in Sumerian records about 2350 B.C., under the name of *lu-ub-sar* (Arabic *lobia*), may have been a reference to the cowpea (Herklots 1972: 261).

Wild peas include *Pisum fulvum*, *Pisum elatius* (*P. sativum* ssp. *elatius*), and *Pisum humile* (*P. sativum* ssp. *humile*, = *Pisum syriacum*). Hybrids between *P. sativum* and *P. elatius* or *P. humile* are usually fertile, whereas hybrids between the cultivated pea and *P. fulvum* are fertile only when *P. fulvum* is the pollen parent (Ladizinsky 1989: 374–7). Zohary (1989a: 363–4) reports that similarities in chromosome structure between cultivated peas and *P. humile* populations in Turkey and the Golan Heights point to populations of *humile* having a particular chromosomal

configuration (a reciprocal translocation) as the “direct” ancestral stock of the cultivated peas.

Peas were present in the early Neolithic preceramic farming villages (7500 to 6000 B.C.) of the Near East, and in large amounts with cultivated wheats and barleys by 5850 to 5600 B.C. They appear to have been associated with the spread of Neolithic agriculture into Europe, again together with wheat and barley (Zohary and Hopf 1988: 96–8). In central Germany, carbonized pea seeds with intact seed coats have been found dating from 4400 to 4200 B.C., and these show the smooth surface characteristic of the domestic crop. In Eastern Europe and Switzerland, pea remains have been found in late Neolithic or Bronze Age sites (Zohary and Hopf 1988: 97–8).

The date at which the pea entered China is not known, but along with the broad bean, it was called *bu-tou* or “Persian bean,” which suggests an origin by way of the Silk Road in historic times (Simoons 1991: 74–5).

Vicia faba

Ladizinsky (1975) has investigated the genetic relationships among the broad bean and its wild relatives in the section *Faba* of the genus *Vicia*. He concluded that none of the populations of wild species, *Vicia narbonensis*, *Vicia galilaea*, or *Vicia baeniscyamus*, collected in Israel, can be considered the ancestor of the fava bean. Differences in the crossability, chromosome form, and chromosome number separate the broad bean from these relatives, as do differences in ranges of seed size, presence of tendrils, and epidermal hairs on the pods. Ladizinsky was careful to note that these characters may have evolved under domestication, but given the absence of evidence for derivation of the fava bean from its known relatives, he concluded that the place of origin of the broad bean could not have been in the Middle East.

Ladizinsky further drew attention to what he regarded as a paucity of broad bean remains only “circumstantial[ly]” identified in Neolithic sites of the Middle East and was led by these lines of evidence to conclude that the origin of *V. faba* had to be outside of this region. He further speculated that the occurrence of a self-pollinating wild form of the species in Afghanistan and areas adjacent might be the region in which the now cross-pollinating broad bean originated. Zohary (1977), however, interpreted the archaeological record quite differently, asserting that the distribution pattern of archaeological broad bean remains points to its domestication by the fourth or fifth millennium B.C. in the Mediterranean Basin.

The dates for the introduction of broad beans into China are not secure – perhaps the second century B.C. or later – but it has become an important crop in “many mountainous, remote or rainy parts of China at the present time especially in western China” (E. N. Anderson, cited by Simoons 1991: 75). The earliest introduction of broad beans into North America

appears to have taken place in the early seventeenth century when Captain Bartholomew Gosnold, who explored the coast of New England, planted them on the Elizabeth Islands off the south shore of Massachusetts (Hedrick 1972: 594). Less than 30 years later, records of the provisioning and outfitting of the supply ship for New Plymouth list "benes" (and "pease") along with other seeds to be sent to the Massachusetts colony (Pulsifer 1861: 24).

Conclusion

The domestication of food grain legumes in the Americas and in the Mediterranean-West Asian region reveals important parallels. Both groups suppressed, at least in some varieties, the tendency to vine or grow rampantly. The seeds of both have markedly increased in size over their wild counterparts. Both groups have suppressed pod dehiscence, projectile seed dissemination, and seed dormancy, and in both regions, the seeds have functioned primarily as protein sources in predominantly cereal grain diets. Studies in genetics, molecular structure, and archaeology have contributed to an understanding of the origins of species and races within species. Nonetheless, uncertainties over important aspects of the origins and evolution under domestication remain, and are the subject of active multidisciplinary research.

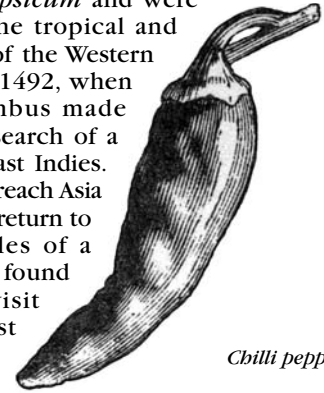
Lawrence Kaplan

Bibliography

- Aykroyd, W. R., and J. Doughty. 1964. *Legumes in human nutrition*. FAO. Rome.
- Beals, R. L. 1932. The comparative ethnology of northern Mexico before 1750. *Ibero-Americana* 2: 93-225.
- Bendremer, J. C. M., E. A. Kellogg, and T. B. Lary. 1991. A grass-lined maize storage pit and early maize horticulture in central Connecticut. *North American Archaeologist* 12: 325-49.
- Blumler, M. A. 1991. Modeling the origin of legume domestication and cultivation. *Economic Botany* 45: 243-50.
- Brooks, R. H., L. Kaplan, H. C. Cutler, and T. H. Whitaker. 1962. Plant material from a cave on the Rio Zape, Durango, Mexico. *American Antiquity* 27: 356-69.
- Butler, A. 1989. Cryptic anatomical characters as evidence of early cultivation in the grain legumes. In *Foraging and farming, the evolution of plant exploitation*, ed. David R. Harris and Gordon C. Hillman, 390-407. London.
- Candolle, Alphonse de. [1886] 1964. *Origin of cultivated plants*. Second edition. New York.
- Castetter, E. F., and W. H. Bell. 1942. *Pima and Papago Indian agriculture*. Albuquerque, N. Mex.
- Darby, William J., Paul Ghalioungui, and Louisa Grivetti. 1977. *Food: The gift of Osiris*. 2 vols. London.
- Delgado Salinas, Alfonso. 1988. Otra interpretación en torno a la domesticación de *Phaseolus*. In *Estudios sobre las revoluciones neolítica y urbana*, ed. L. Manzanilla, 167-74. Mexico.
- Duke, J. A. 1981. *Handbook of legumes of world economic importance*. New York.
- Dunn, O., and J. E. Kelley, Jr. 1989. *The Diario of Christopher Columbus's first voyage to America 1492-1493*. Norman, Okla.
- Franquemont, C., E. Franquemont, W. Davis, et al. 1990. *The ethnobotany of Chinchero, an Andean community of southern Peru*. Field Museum of Natural History, Chicago, *Fieldiana*, Publication No. 1408.
- Freeman, G. F. 1912. Southwestern beans and teparies. *Arizona Agricultural Experiment Station Bulletin* 68. Phoenix.
- Gepts, P. A. 1988. *Genetic resources of Phaseolus beans*. Dordrecht, Netherlands.
- Hansen, J. 1992. Franchthi cave and the beginnings of agriculture in Greece and the Aegean. In *Préhistoire de l'agriculture: Nouvelles approches expérimentales et ethnographiques*. Paris.
- Hedrick, U. P. 1931. *The vegetables of New York*, Vol. 1, Part 2, *Beans of New York*. Albany, N.Y.
- ed. [1919] 1972. *Sturtevant's notes on edible plants*. New York.
- Heiser, C. B., Jr. 1965. Cultivated plants and cultural diffusion in nuclear America. *American Anthropologist* 67: 930-49.
- Hendry, G. W. 1934. Bean cultivation in California. *California Agricultural Experiment Station Bulletin* 294: 288-321.
- Herklots, G. A. C. 1972. *Vegetables in Southeast Asia*. London.
- Kaplan, L. 1967. Archeological *Phaseolus* from Tehuacan. In *The prehistory of the Tehuacan Valley*, ed. Douglas S. Byers, 201-12. Austin, Tex.
1970. Plant remains from the Blain Site. In *Blain Village and the Fort Ancient tradition in Ohio*, ed. Olaf H. Prufer and Orrin C. Shane, III, 227-31. Kent, Ohio.
1981. What is the origin of the common bean? *Economic Botany* 35: 240-54.
1986. Pre-ceramic *Phaseolus* from Guila' Naquitz. In *Guila' Naquitz: Archaic foraging and early agriculture in Oaxaca*, ed. Kent Flannery, 281-4. Orlando, Fla.
1994. Accelerator Mass Spectrometry date and the antiquity of *Phaseolus* cultivation. *Annual Report of the Bean Improvement Cooperative* 37: 131-2.
1995. Accelerator dates and the prehistory of *Phaseolus*. Paper contributed to the Annual Meeting, Society for American Archaeology, May 5, 1995, Minneapolis, Minn.
- Kaplan, L., T. F. Lynch, and C. E. Smith, Jr. 1973. Early cultivated beans (*Phaseolus vulgaris*) from an intermontane Peruvian valley. *Science* 179: 76-7.
- Kaplan, L., and R. S. MacNeish. 1960. Prehistoric bean remains from caves in the Ocampo region of Tamaulipas, Mexico. *Botanical Museum Leaflets* (Harvard University) 19: 33-56.
- Kislev, M. E. 1985. Early Neolithic horsebean from Yiftah'el, Israel. *Science* 228: 319-20.
- Ladizinsky, G. 1975. On the origin of the broad bean, *Vicia faba* L. *Israel Journal of Botany* 24: 80-8.
1979. Seed dispersal in relation to the domestication of Middle East legumes. *Economic Botany* 33: 284-9.
1987. Pulse domestication before cultivation. *Economic Botany* 41: 60-5.
1989. Origin and domestication of the southwest Asian grain legumes. In *Foraging and farming, the evolution of plant exploitation*, ed. David R. Harris and Gordon C. Hillman, 374-89. London.
1993. Lentil domestication: On the quality of evidence and arguments. *Economic Botany* 47: 60-4.

- Ladizinsky, G., D. Braun, D. Goshen, and F. J. Muehlbauer. 1984. The biological species of the genus *Lens*. *Botanical Gazette* 145: 253–61.
- Larco Hoyle, Rafael. 1943. La escritura Mochica sobre palares. *Revista Geográfica Americana* 20: 1–36.
- Lynch, T. F., R. Gillespie, J. A. Gowlette, and R. E. M. Hedges. 1985. Chronology of Guitarrero Cave, Peru. *Science* 229: 864–7.
- Oxford English Dictionary*. 1971. Compact edition.
- Pulsifer, D., ed. 1861. *Records of the Colony of New Plymouth in New England, 1623–1682*, Vol. 1. Boston, Mass.
- Purseglove, J. W. 1968. *Tropical crops: Dicotyledons*. 2 vols. New York.
- Schmit, V., and D. G. Debouck. 1991. Observations on the origin of *Phaseolus polyanthus* Greenman. *Economic Botany* 45: 345–64.
- Simoons, F. J. 1991. *Food in China, a cultural and historical inquiry*. Boca Raton, Fla.
- Singh, U. and B. Singh. 1992. Tropical grain legumes as important human foods. *Economic Botany* 46: 310–21.
- Smartt, J. 1990. *Grain legumes, evolution and genetic resources*. Cambridge and New York.
- Sokolov, Raymond. 1984. Broad bean universe. *Natural History* 12: 84–7.
- Spencer, R. F., J. D. Jennings, et al. 1965. *The Native Americans: Prehistory and ethnology of the North American Indians*. New York.
- Strickberger, M. W. 1985. *Genetics*. Third edition. New York.
- Tournefort, J. P. 1730. *The compleat herbal, or the botanical institutions of Monsr. Tournefort, Chief Botanist to the late French King*. . . 2 vols. London.
- Vavilov, N. I. 1992. *Origin and geography of cultivated plants*, ed. V. F. Dorofeyev, trans. Doris Löve. Cambridge.
- Webster's third new international dictionary of the English language*. Unabridged. Springfield, Mass.
- Wills, W. H. 1988. *Early prehistoric agriculture in the American Southwest*. Santa Fe, N. Mex.
- Zohary, D. 1977. Comments on the origin of cultivated broad bean. *Israel Journal of Botany* 26: 39–40.
- 1989a. Domestication of the southwest Asian Neolithic crop assemblage of cereals, pulses, and flax: The evidence from living plants. In *Foraging and farming, the evolution of plant exploitation*, ed. David R. Harris and Gordon C. Hillman, 358–73. London.
- 1989b. Pulse domestication and cereal domestication: How different are they? *Economic Botany* 43: 31–4.
- Zohary, D., and M. Hopf. 1988. *Domestication of plants in the Old World*. Oxford.

ing to the genus *Capsicum* and were unknown outside the tropical and subtropical regions of the Western Hemisphere before 1492, when Christopher Columbus made his epic voyage in search of a short route to the East Indies. Although he did not reach Asia and its spices, he did return to Spain with examples of a new, pungent spice found during his first visit to the eastern coast of the Caribbean island of Hispaniola



Chilli pepper

(now the Dominican Republic and Republic of Haiti). Today capsicums are not only consumed as a spice, condiment, and vegetable, but are also used in medicines, as coloring agents, for landscape and decorative design, and as ornamental objects.

History

For the peoples of the Old World, the history of capsicums began at the end of the fifteenth century, when Columbus brought some specimens of a red-fruited plant from the New World back to his sovereigns (Morison 1963: 216; Anghiera 1964: 225). However, the fruits were not new to humankind. When nonagricultural Mongoloid peoples, who had begun migrating across the Bering Strait during the last Ice Age, reached the subtropical and tropical zones of their new world, they found capsicums that had already become rather widespread. They had been carried to other regions by natural dispersal agents – principally birds – from their nuclear area south of both the wet forests of Amazonia and the semiarid cerado of central Brazil (Pickersgill 1984: 110). Plant remains and depictions of chillies on artifacts provide archaeological evidence of the use and probable cultivation of these wild capsicums by humans as early as 5000 B.C. By 1492, Native Americans had domesticated (genetically altered) at least four species (MacNeish 1967; Heiser 1976: 266; Pickersgill 1984: 113). No others have subsequently been domesticated.

In the West Indies, Columbus found several different capsicums being cultivated by the Arawak Indians, who had migrated from northeastern South America to the Caribbean Islands during a 1,200 year period beginning about 1000 B.C. (Means 1935; Anghiera 1964; Watts 1987). These migrants traveled by way of Trinidad and the Lesser Antilles, bringing with them a tropical capsicum that had been domesticated in their homeland. They also brought a word similar to *ají* by which the plant was, and still is, known in the West Indies and throughout its native South American habitat (Heiser 1969). Later, a second species reached the West Indies from Mesoamerica along with other food plants, such as maize (corn), beans, and squash (Sauer

II.C.4 🌶️ Chilli Peppers

Chilli peppers are eaten as a spice and as a condiment by more than one-quarter of the earth's inhabitants each day. Many more eat them with varying regularity – and the rate of consumption is growing. Although the chilli pepper is the most used spice and condiment in the world, its monetary value in the spice trade is not indicative of this importance because it is readily cultivated by many of its consumers.

Peppers are the fruit of perennial shrubs belong-

1966: 54). It was this, more climatically adaptable pepper species that went forth, bearing the native Nahuatl name *chilli*, to set the cuisines of the Old World tropics afire (Andrews 1993a, 1993b).

The conquest of Mexico and, later, the mastery of Peru also revealed pepper varieties more suited climatically to cultivation in the temperate areas of Europe and the Middle East. And within 50 years after the first capsicum peppers reached the Iberian Peninsula from the West Indies, American chilli peppers were being grown on all coasts of Africa and in India, monsoon Asia, southwestern China, the Middle East, the Balkans, central Europe, and Italy (Andrews 1993a).

The first European depictions of peppers date from 1542, when a German herbal by Leonhart Fuchs described and illustrated several types of pepper plants considered at that time to be native to India. Interestingly, however, it was not the Spanish who were responsible for the early diffusion of New World food plants. Rather, it was the Portuguese, aided by local traders following long-used trade routes, who spread American plants throughout the Old World with almost unbelievable rapidity (Boxer 1969a).

Unfortunately, documentation for the routes that chilli peppers followed from the Americas is not as plentiful as that for other New World economic plants such as maize, tobacco, sweet potatoes, manioc (cassava), beans, and tomatoes. However, it is highly probable that capsicums accompanied the better-documented Mesoamerican food complex of corn, beans, and squash, as peppers have been closely associated with these plants throughout history. The Portuguese, for example, acquired corn at the beginning of the sixteenth century and encouraged its cultivation on the west coast of Africa (Jeffreys 1975: 35). From West Africa the American foods, including capsicums, went to the east coast of Africa and then to India on trading ships traveling between Lisbon and Goa on the Malabar Coast of western India (Boxer 1984).

The fiery new spice was readily accepted by the natives of Africa and India, who were long-accustomed to food highly seasoned with spices, such as the African melegueta pepper (*Aframomum melegueta*, also known as "grains of paradise"), the Indian black pepper (*Piper nigrum*), and ginger (*Zingiber officinale*). In fact, because the plants produced by the abundant, easily stored seeds were much easier to cultivate than the native spices, *Capsicum* became a less expensive addition to the daily diet and was soon widely available to all – rich and poor alike. Thus, within a scant 50 years after 1492, three varieties of capsicums were being grown and exported along the Malabar Coast of India (Purseglove 1968; Watt 1972).

From India, chilli peppers traveled (along with the other spices that were disseminated) not only along the Portuguese route back around Africa to Europe but also over ancient trade routes that led either to

Europe via the Middle East or to monsoon Asia (L'obel 1576). In the latter case, if the Portuguese had not carried chilli peppers to Southeast Asia and Japan, the new spice would have been spread by Arabic, Gujarati, Chinese, Malaysian, Vietnamese, and Javanese traders as they traded traditional wares throughout their worlds. And, after Portuguese introduction, both birds and humans carried the peppers inland. Certainly, birds are most adept at carrying pepper seeds from island to island and to inaccessible inland areas (Ridley 1930; Procter 1968).

In the Szechuan and Hunan provinces in China, where many New World foods were established within the lifetime of the Spanish conquistadors, there were no roads leading from the coast. Nonetheless, American foods were known there by the middle of the sixteenth century, having reached these regions via caravan routes from the Ganges River through Burma and across western China (Ho 1955). The cuisines of southwestern Szechuan and Hunan still employ more chilli peppers than any other area in China.

Despite a European "discovery" of the Americas, chilli peppers diffused throughout Europe in circuitous fashion. Following the fall of Granada in 1492, the Spaniards established dominance over the western Mediterranean while the Ottoman Turks succeeded in installing themselves as the controlling power in northern Africa, Egypt, Arabia, the Balkans, the Middle East, and the eastern Mediterranean. The result was that the Mediterranean became, in reality, two separate seas divided by Italy, Malta, and Sicily, with little or no trade or contact between the eastern and western sections (Braudel 1976).

Venice was the center of the spice and Oriental trade of central Europe, and Venice depended on the Ottoman Turks for goods from the fabled Orient. From central Europe the trade went to Antwerp and the rest of Europe, although Antwerp also received Far Eastern goods from the Portuguese via India, Africa, and Lisbon. It was along these avenues that chilli peppers traveled into much of Europe. They were in Italy by 1535 (Oviedo 1950), Germany by 1542 (Fuchs 1543), England before 1538 (Turner 1965), the Balkans before 1569 (Halasz 1963), and Moravia by 1585 (L'Escluse 1611). But except in the Balkans and Turkey, Europeans did not make much use of chilli peppers until the Napoleonic blockade cut off their supply of spices and they turned to Balkan paprika as a substitute. Prior to that, Europeans had mainly grown capsicums in containers as ornamentals.

Well into the nineteenth century, most Europeans continued to believe that peppers were native to India and the Orient until botanist Alphonse de Candolle produced convincing linguistic evidence for the American origin of the genus *Capsicum* (Candolle 1852). In addition, during the 500 years since their discovery, chillies have become an established crop in the Old World tropics and are such a vital part of the

cuisines that many in these regions are only now beginning to accept an American origin of the spice that is such an integral part of their daily lives.

It was only after the Portuguese had carried capsicums and other American plants to Africa, Asia, the Middle East, and Europe that the Spaniards played a significant role in the movement of New World crops to places other than Spain, Italy, and, perhaps, Western Europe. This began toward the end of the sixteenth century with the Manila-Acapulco galleon traffic which effected the transfer of numerous plants, as well as goods, between Mexico and the Orient (Schurz 1939).

Moreover, in North America, at approximately the same time that the Manila galleon trade was launched, the Spaniards founded the presidios of Saint Augustine, Florida (1565), and Santa Fe, New Mexico (1598). These settlements initiated Caribbean-Florida and Mexico-American Southwest exchanges of plants long before other Europeans began colonizing the east coast of North America. Interestingly, however, seventeenth-century English colonists introduced peppers from England via Bermuda to their eastern North American possessions (Laufer 1929: 242).

The Names of Chilli Peppers

That Columbus had not reached the Orient did not discourage him from calling the Caribbean Islands the "Indies," the natives "Indians," and the chilli pepper *pimiento* after the completely unrelated black pepper - *pimenta* - that he sought in the East.

The indigenous Arawaks called the fruit *axí*, which was the South American name they brought with them when they migrated north to the Antilles. Although the Spaniards transliterated this to *ají* (*ajé*, *agí*), they never adopted the Arawak word, either in the West Indies or in North America.

Nonetheless, in the Dominican Republic and a few other places in the Caribbean, and in much of South America, the pungent varieties are still called *ají*. *Uchu* and *buayca* are other ancient words used for capsicums by some Amerindian groups in the Andean area. In Spain, American peppers are called *pimiento* or *pimentón* (depending on the size) after *pimenta* (black pepper from India). In Italy, they are called *peperone*, in France, *piment*, and in the Balkans, *paprika*.

In Mexico, however, the Nahuatl-speaking natives called their fiery fruit *chilli*. The Nahuatl stem *chil* refers to the chilli plant. It also means "red." The original Spanish spelling was *chilli*, first used in print by Francisco Hernández (1514-78), the earliest European to collect plants systematically in the New World. But although in his writings (published in 1615) he interpreted the Nahuatl name for capsicums as chilli, that Spanish spelling was later changed to *chile* by Spanish-speaking people in Mexico. To the generic word "chilli" were added the terms that

described particular chilli cultivars (two examples are *Tonalchilli*, "chilli of the sun or summer," and *Chiltecpin*, "flea chilli"). In Mexico today, the word *chile* refers to both pungent and sweet types and is used, in the Nahuatl style, combined with a descriptive adjective, such as *chile colorado* ("red chilli"), or with a word that indicates the place of origin, such as *chile poblano* ("chilli from Puebla"). The same Mexican variety can have different names in different geographic regions, in various stages of maturity, and in the dried state.

The Portuguese language uses *pimenta* for capsicums and qualifies the various types - *pimenta-da-caiena* (cayenne pepper), *pimenta-da-malagueta* (red pepper), *pimenta-do-reino* (black pepper), and *pimenta-da-jamaica* (allspice). *Pimentão* can mean pimiento, red pepper, or just pepper. *Ají* and *chile* are not found in Portuguese dictionaries, and apparently the Portuguese did not carry those words with them in their travels.

The Dutch and the English were probably responsible for introducing the current capsicum names to the eastern part of the Old World. In Australia, India, Indonesia, and Southeast Asia in general, the term "chilli" ("chillies") or, sometimes, "chilly," is used by English speakers for the pungent types, whereas the mild ones are called capsicums. Each Far Eastern language has its own word for chillies - *prik* in Thai and *mirch* in Hindi, to name but two.

It is in the United States that the greatest confusion exists. Both the Anglicized spelling, "chili" (chillies), and the Spanish *chile* (*chiles*) are used by some for the fruits of the *Capsicum* plant, but chili is also used as a short form of *chili con carne*, a variously concocted mixture of meat and chillies. The *Oxford English Dictionary* designates "chilli" (after the Nahuatl) as the primary usage, calling the Spanish *chile* and the English chili both variants. *Webster's New International Dictionary*, however, prefers "chili" followed by the Spanish *chile* and the Nahuatl *chilli*. But "chilli" is the term most often used by English-speaking people outside the United States, and it is the spelling preferred by the International Board for Plant Genetic Resources (IBPGR) of the Food and Agriculture Organization of the United Nations (FAO).

Origin

It is difficult to determine exactly where the genus *Capsicum* originated because the nature of that genus is still not fully understood (Eshbaugh 1980). If the genus remains limited to taxa producing the pungent capsaicin, then the center of diversity occurs in an area from Bolivia to southwestern Brazil. But if the genus includes nonpungent taxa, a second center of diversity would center in Mesoamerica. Nonetheless, it is certain that the ancestor of all of the domesticates originated in tropical South America.

There are definite indications that *Capsicum*

annuum originally was domesticated in Mesoamerica and *Capsicum chinense* in tropical northern Amazonia. *Capsicum pubescens* and *C. baccatum* seem to be more commonplace in the Andean and central regions of South America. Thus, the first two species were those encountered by the first Europeans, whereas the other two species were not found until later and are just now becoming known outside their South American home.

Diagnostic Descriptions

The genus *Capsicum* is of the family Solanaceae, which includes such plants as the potato, tomato, eggplant, petunia, and tobacco. The genus was first described in 1700, but that description has become so outdated as to be worthless. The taxonomy of the genus *Capsicum* is in a state of transition, and the taxa finally included may change if the description is expanded to encompass taxa with common traits but nonpungent fruits (Eshbaugh, personal communication).

Currently, the genus consists of at least 20 species, many of which are consumed by humans. Four of the species have been domesticated and two others are extensively cultivated. It is those six species, belonging to three separate genetic lineages, that are of concern to human nutrition.

Capsicum pubescens Ruiz and Pavón

The domesticated *C. pubescens* is the most distinctive species in the genus. The flowers have comparatively large purple or white (infused with purple) corollas that are solitary and erect at each node. Those blossoms, along with the wavy, dark brownish black seeds, are unique among the capsicums. This extremely pungent chilli was domesticated in the Andean region of South America, where it is commonly called *rocoto*, and it is still practically unknown in other parts of the world because it requires cool but frost-free growing conditions and a long growing season at relatively high elevations. Its many varieties include none that are sweet. The fleshy nature of the fruit causes rapid deterioration when mature, and, consequently, it neither travels nor stores well.

Capsicum baccatum var. *pendulum* (Willdenow) Eshbaugh

Capsicum baccatum var. *pendulum* is recognized by a flower with a cream-colored corolla marked with greenish-gold blotches near the base of each petal and anthers that are whitish-yellow to brown. It is solitary at each node. Although it is quite variable, the typical fruit is elongate with cream-colored seeds. It is indigenous to the lowlands and mid-elevations of Bolivia and neighboring areas. In much of South America, where all pungent peppers are called *ají*, *C. baccatum* is the "Andean *ají*" (Ruskin 1990: 197). Until recently, it has been little known outside South America. It is only in this species and the common annual

pepper that nonpungent cultivars are known (Ruskin 1990: 198).

Capsicum annum var. *annuum* Linné

The flowers of *C. var. annum* are solitary at each node (occasionally two or more). The corolla is milky white and the anthers are purple. The variform fruit usually has firm flesh and straw-colored seeds. The pungent and nonpungent cultivars of this Mesoamerican domesticate now dominate the commercial pepper market throughout the world. A relationship between *C. annum*, *C. chinense*, and *Capsicum frutescens* has caused the three to be known as the "*C. annum* complex." This relationship, however, creates a taxonomic predicament, because some authorities still recognize the first two as distinct but have difficulty determining where *C. frutescens* fits into the picture.

Capsicum annum var. *glabrisculum*

Capsicum annum var. *glabrisculum* is a semiwild species known as bird pepper. This highly variable, tiny, erect, usually red pepper is cultivated commercially in the area around Sonora, Mexico, and seems to be in the process of domestication. It has a distinct flavor and high pungency and is avidly consumed throughout its natural range, which extends through the southernmost parts of the United States to Colombia. Birds are also keen consumers. These chillies, which have many vernacular names and almost as many synonyms (*Capsicum aviculare* is the most common), sell for 10 times the price of cultivated green bell peppers.

Capsicum chinense Jacquin

There are two or more small, white-to-greenish-white flowers with purple anthers per node of *C. chinense*, often hanging in clusters. The fruit is variform, with cream-colored seeds that tend to require a longer germination period than *C. annum*. *C. chinense* was domesticated in the lowland jungle of the western Amazon River basin and was carried to the islands of the Caribbean before 1492. It has diffused throughout the world but to a much lesser degree than *C. annum*, probably because it does not store or dry well. Nonetheless, it is becoming ever more widely appreciated by cooks and gardeners for its pungency, aroma, and unique flavor, and ever more important in medical, pharmaceutical, and food-industry applications because of its high capsaicin content. Although this morphologically distinct pepper is still considered to be a part of the *C. annum* complex, there are those who question its position in the genus on genetic grounds.

Capsicum frutescens Linné

Some authors no longer list the semiwild *C. frutescens* as a sustainable species. Although it was once considered to be a member of the *C. annum* com-

plex, which included three white-flowered species thought to have a mutual ancestor, scholars now have considerable doubt as to the position of the first two in the genus. The small greenish-white flowers of *C. frutescens* have purple anthers. The small fruit with cream-colored seed is always erect, never sweet, and two or more occur at each node. The tabasco pepper is the only variety of this species known to have been cultivated commercially, and this activity has been limited to the Western Hemisphere.

Geographic Distribution

Following the arrival of the Europeans in the Western Hemisphere, the tropical perennial capsicum spread rapidly. It quickly became pantropic and the dominant spice and condiment in the tropical and subtropical areas of the world. In addition, it is an important green vegetable throughout the temperate regions, where it is grown as an annual. Concentrated breeding studies are producing *Capsicum* varieties that can be cultivated in environments quite different from the tropical home of the original.

Biology

Nutritional Considerations

Capsicums have a lot to recommend them nutritionally. By weight, they contain more vitamin A than any other food plant, and they are also a good source of the B vitamins. When eaten raw, capsicums are superior to citrus in providing vitamin C, although their production of vitamin C diminishes with maturity and drying and (as in all plant foods) is destroyed by exposure to oxygen. By contrast, vitamin A increases as peppers mature and dry and is not affected by exposure to oxygen.

Capsicums also contain significant amounts of magnesium and iron. Chillies, of course, are not eaten in large quantities, but even small amounts are important in cases where traditional diets provide only marginal supplies of vitamins and minerals.

The Pungent Principle

A unique group of mouth-warming, amide-type alkaloids, containing a small vanilloid structural component, is responsible for the burning sensation associated with capsicums by acting directly on the pain receptors in the mouth and throat. This vanilloid element is present in other pungent plants used for spices, like ginger and black pepper. Birds and certain other creatures, such as snails and frogs, do not have specific neuroreceptors for pungent vanilloid compounds as do humans and other mammals; consequently, their contact with capsaicinoids has no adverse effects (Nabhan 1985).

The vanillyl amide compounds or capsaicinoids (abbreviated CAPS) in *Capsicum* are predominantly (about 69 percent) capsaicin (C). Dihydrocapsaicin

(DHC) (22 percent), nordihydrocapsaicin (NDHC) (7 percent), homocapsaicin (HC) (1 percent), and homodihydrocapsaicin (HDHC) (1 percent) account for most of the remainder (Masada et al. 1971; Trease and Evans 1983). The primary heat contributors are C and DHC, but the delayed action of HDHC is the most irritating and difficult to quell (Mathew et al. 1971).

Three of these capsaicinoid components cause the sensation of "rapid bite" at the back of the palate and throat, and two others cause a long, low-intensity bite on the tongue and the middle palate. Differences in the proportions of these compounds may account for the characteristic "burns" of the different types of capsicum cultivars (McGee 1984: 12; Govindarajan 1986).

In both sweet and pungent capsicums, the major part of the organs secreting these pungent alkaloids is localized in the placenta, to which the seeds are attached, along with dissepiment (veins or cross walls) (Heiser and Smith 1953). The seeds contain only a low concentration of CAPS. The capsaicin content is influenced by the growing conditions of the plant and the age of the fruit and is possibly variety-specific (Govindarajan 1986: 336-8). Dry, stressful conditions will increase the amount of CAPS. Beginning about the eleventh day of fruit development, the CAPS content increases, becoming detectable when the fruit is about four weeks old. It reaches its peak just before maturity, then drops somewhat in the ripening stage (Govindarajan 1985). Sun-drying generally reduces the CAPS content, whereas the highest retention of CAPS is obtained when the fruits are air-dried with minimum exposure to sunlight.

Capsaicin is hard to detect by chemical tests. It has virtually no odor or flavor, but a drop of a solution containing one part in 100,000 causes a persistent burning on the tongue (Nelson 1910). The original Scoville Organoleptic Test has largely been replaced by the use of high-pressure liquid chromatography (HPLC), a highly reproducible technique for quantifying capsaicinoids in capsicum products. However, the results apply solely to the fruit tested, and therefore they are considered only as a general guide (Todd, Bensinger, and Biftu 1977). Capsaicin is eight times more pungent than the piperine in black pepper. But unlike black pepper, which inhibits all tastes, CAPS obstructs only the perception of sour and bitter; it does not impair the discernment of other gustatory characteristics of food.

Capsaicin activates the defensive and digestive systems by acting as an irritant to the oral and gastrointestinal membranes (Viranuvatti et al. 1972). That irritation increases the flow of saliva and gastric acids and also stimulates the appetite. These functions work together to aid the digestion of food. The increased saliva helps ease the passage of food through the mouth to the stomach, where it is mixed with the activated gastric juice (Solanke 1973). Ingesting CAPS also causes the neck, face, and front of the chest to sweat in a reflexive response to the burning

in the mouth (Lee 1954). Very little CAPS is absorbed as it passes through the digestive tract (Diehl and Bauer 1978).

Capsaicin is not water soluble, but the addition of a small amount of chlorine or ammonia will ionize the CAPS compound, changing it into a soluble salt (Andrews 1984: 127) that can be used to rinse CAPS from the skin. Like many organic compounds, CAPS is soluble in alcohol. Oral burning can be relieved by lipoproteins, such as casein, that remove CAPS by breaking the bond it has formed with the pain receptors in the mouth (Henkin 1991). Milk and yoghurt are the most readily available sources of the casein. Because casein, and not fat, removes capsaicin, butter and cheese will not have the same effect as milk.

Studies of CAPS and its relationship to substance P, a neuropeptide that sends the message of pain to our brains, have led investigators to conclude that CAPS has the capacity to deplete nerves of their supply of substance P, thereby preventing the transmission of such messages (Rozin 1990).

Consequently, CAPS is now used to treat the pain associated with shingles, rheumatoid arthritis, and "phantom-limb" pain. It may prove to be a non-addictive alternative to the habit-forming drugs used to control pain from other causes. It does not act on other sensory receptors, such as those for taste and smell, but is specific to pain receptors. Such specificity is becoming a valuable aid to medical research.

Aroma, Flavor, and Color

The flavor compound of capsicums is located in the outer wall of the fruit (pericarp): Very little is found in the placenta and cross wall and essentially none in the seeds (Figure II.C.4.1). Color and flavor go hand in hand because the flavoring principle appears to be associated with the carotenoid pigment: Strong color and strong flavor are linked. *Capsicum pubescens* (rocoto) and the varieties of *C. chinense* are more aromatic and have a decidedly different flavor from those of *C. annuum* var. *annuum*. The carotenoid pigments responsible for the color in capsicums make peppers commercially important worldwide as natural dyes in food and drug products. Red capsanthin is the most important pigment. All capsicums will change color from green to other hues – red, brown, yellow, orange, purple, and ripe green – as they mature.

Taste and smell are separate perceptions. Several aroma compounds produce the fragrance. The taste buds on the tongue can discern certain flavors at dilutions up to one part in two million, but odors can be detected at a dilution of one part in one billion. The more delicate flavors of foods are recognized as aromas in the nasal cavity adjacent to the mouth.

Cultivation Requirements

Peppers are best transplanted and not planted directly into the soil outdoors. The seeds should be started in greenhouse benches, flats, or hotbeds at

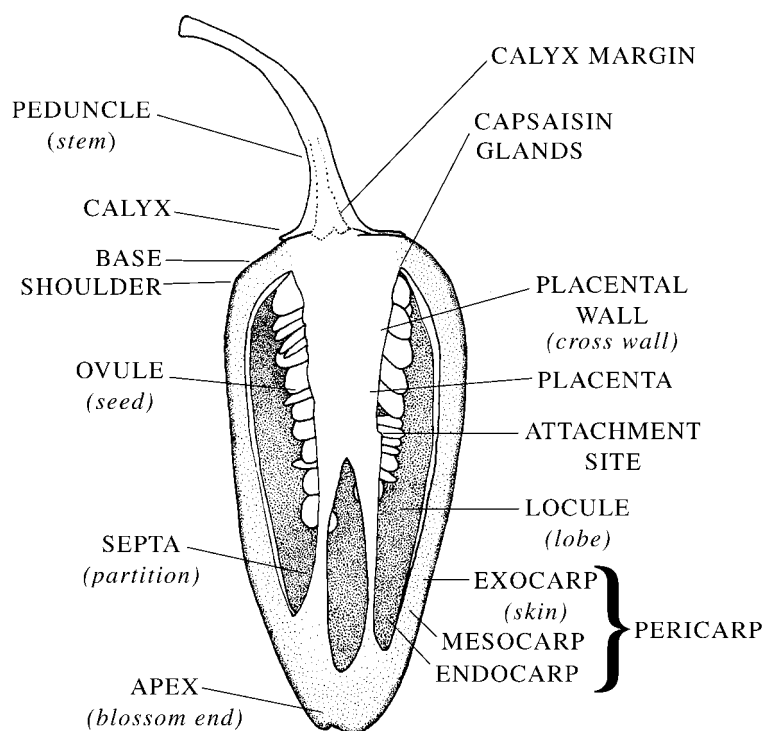


Figure II.C.4.1. Cross-section of a pepper.
(Adapted from Andrews 1995.)

least six weeks before the first frost-free date. They ought to be sown as thinly as possible on a sterile medium and covered no deeper than the thickness of the seed. It is best to water them from the top, with care taken to not dislodge the seed. The seed or seedlings should never be permitted to dry or wilt from the time they are sown until they are transplanted and well started. Germination will require 12 to 21 days at a constant temperature of 21° C for *C. annuum* var. *annuum*, and longer for the other species.

When the true leaves are well formed, one may transplant the seedlings into containers or flats, containing equal parts peat, sand, and loam, and grow them at 21° C. After the plants attain a height of 12 to 15 centimeters (cm), and all danger of frost is past, they can be planted (deeply) in friable soil that is not below 13° C. The plants should be spaced 30 cm apart in rows 38 to 76 cm apart. Peppers require full sun and well-drained soil. They are warm-season plants that do better in a moderate climate, with the optimum temperature for good yields between 18.5° C and 26.5° C during fruit setting (Andrews 1984).

Economic and Other Uses

Perhaps no other cultivated economic plants have fruits with so many shapes, colors, and uses over such a widespread area of the earth as do those belonging to the genus *Capsicum*.

Before World War II, capsicums were eaten daily by one-fourth of the world's population, primarily in the pantropical belt and Korea. Since that time, their consumption as a condiment, spice, and vegetable has continued to increase annually and is no longer limited to the tropical and subtropical areas. Some of the more common food products made with chillies are curry powder, cayenne pepper, crushed red pepper, dried whole peppers, chili powder, paprika, pepper sauce, pickled and processed peppers, pimento, and salsa picante. In 1992, the monetary value of sales of salsa picante, a bottled sauce of Mexican origin made with chillies, onions, and tomatoes, overtook that of tomato catsup in the United States.

However, the use of capsicums goes beyond that of food. The florist and landscape industries have discovered the ornamental qualities of pepper plants to be of considerable value, and designers of tableware, home decorations, fabrics, and paper goods find them to be a popular decorative motif. The medical profession has discovered that certain folk-medicine practices employing chillies, some of which are prehistoric in origin, have merit. Capsaicin, the pungent alkaloid unique to capsicums, is being utilized in modern medicine to treat pain, respiratory disorders, shingles, toothache, and arthritis, and research into the properties of capsaicinoids continues.

Jean Andrews

Bibliography

- Andrews, J. 1993a. Diffusion of the Mesoamerican food complex to southeastern Europe. *Geographical Review* 83: 194-204.
- 1993b. *Red hot peppers*. New York.
- [1984] 1995. *Peppers: The domesticated capsicums*. Austin, Tex.
- Anghiera, P. M. d'. 1964. *Decadas del Nuevo Mundo, por Pedro Martir de Angleria, primer cronista de Indias*. Mexico City.
- Boxer, C. R. 1969a. *Four centuries of Portuguese expansion: 1415-1825*. Berkeley, Calif.
- 1969b. *The Portuguese seaborne empire, 1415-1825*. London.
1984. *From Lisbon to Goa 1500-1750. Studies in Portuguese maritime enterprise*. London.
1985. *Portuguese conquest and commerce in southern Asia 1500-1750*. London.
- Braudel, F. 1976. *The Mediterranean and the Mediterranean world in the age of Philip II*. 2 vols. New York.
- Candolle, A. P. de. 1852. *Essai. Prodromus* 13: 411-29. Paris.
- Columbus, C. 1971. *Journal of first voyage to America by Christopher Columbus*. Freeport, N.Y.
- Diehl, A. K., and R. L. Bauer. 1978. Jaloproctitis. *New England Journal of Medicine* 229: 1137-8.
- Eshbaugh, W. H. 1968. A nomenclatural note on the genus *Capsicum*. *Taxon* 17: 51-2.
1980. The taxonomy of the genus *Capsicum* (Solanaceae). *Phytologia* 47: 153-66.
1983. The genus *Capsicum* (Solanaceae) in Africa. *Bothalia* 14: 845-8.
1993. Peppers: History and exploitation of a serendipitous new crop. In *Advances in new crops*, ed. J. J. Janick and J. E. Simon. New York.
- Eshbaugh, W. H., S. I. Guttman, and M. J. McLeod. 1983. The origin and evolution of domesticated *Capsicum* species. *Journal of Ethnobiology* 3: 49-54.
- Fuchs, L. 1543. *New Kreuterbuch (De historia stirpium in 1542)*. Basel.
- Govindarajan, V. S. 1985. *Capsicum*: Production, technology, chemistry and quality. History, botany, cultivation and primary processing. *Critical Review of Food Science and Nutrition* 22: 108-75.
1986. *Capsicum*: Production, technology, chemistry and quality. Chemistry of the color, aroma, and pungency stimuli. *Critical Review of Food Science and Nutrition* 24: 244-355.
- Halasz, Z. 1963. *Hungarian paprika through the ages*. Budapest.
- Heiser, C. B., Jr. 1969. *Nightshades: The paradoxical plants*. San Francisco, Calif.
1976. Peppers: *Capsicum* (Solanaceae). In *Evolution of crop plants*, ed. N. W. Simmonds, 265-8. London.
1985. *Of plants and man*. Norman, Okla.
- Heiser, C. B., Jr., and B. Pickersgill. 1975. Names for the bird peppers (*Capsicum* - Solanaceae). *Baileya* 19: 151-6.
- Heiser, C. B., Jr., and P. G. Smith. 1953. The cultivated *Capsicum* peppers. *Economic Botany* 7: 214-27.
- Henkin, R. 1991. Cooling the burn from hot peppers. *Journal of the American Medical Association* 266: 2766.
- Ho, P. T. 1955. The introduction of American food plants into China. *American Anthropologist* 55: 191-201.
- Jacquin, N. J. 1776. *Hortus botanicus vindobonensis*. 3 vols. Vienna.
- Jeffreys, M. D. W. 1975. Pre-Columbian maize in the Old

- World: An examination of Portuguese sources. In *Gastronomy: The anthropology of food and food habits*, ed. M. L. Arnott, 23–66. The Hague.
- Laufer, B. 1929. The American plant migration. *The Scientific Monthly* 28: 235–51.
- Lee, T. S. 1954. Physiological gustatory sweating in a warm climate. *Journal of Physiology* 124: 528–42.
- L'Escluse, C. 1611. *Curae posteriores post mortem*. Antwerp.
- Linnaeus, C. 1753a. *Hortus cliffortianus*. Amsterdam.
- 1753b. *Species plantarum*. Stockholm.
- L'obel, M. 1576. *Plantarum sev stirpium historia*. Antwerp.
- MacNeish, R. S. 1967. A summary of the subsistence. In *The prehistory of the Tebucan Valley*. Vol. 1, *Environment and subsistence*, ed. D. S. Byres, 290–309. Austin, Tex.
- Maga, J. A. 1975. *Capsicum*. In *Critical revisions in food science and nutrition*, 177–99. Cleveland.
- Masada, Y., K. Hashimoto, T. Inoue, and M. Suzui. 1971. Analysis of the pungent principles of *Capsicum annum* by combined gas chromatography. *Journal of Food Science* 36: 858.
- Mathew, A. G., Y. S. Lewis, N. Kirishnamurthy, and E. S. Nambudiri. 1971. Capsaicin. *The Flavor Industry* 2: 691–5.
- McGee, H. 1984. *On food and cooking: The science and lore of the kitchen*. New York.
- McLeod, M. J. S., S. I. Guttman, and W. H. Eshbaugh. 1982. Early evolution of chili peppers (*Capsicum*). *Economic Botany* 36: 361–8.
- Means, P. A. 1935. *The Spanish Main: Focus on envy 1492–1700*. New York.
- Morison, S. E. 1963. *The journals and other documents of the life of Christopher Columbus*. New York.
- Nabhan, G. P. 1985. *Gathering the desert*. Tucson.
- Nelson, E. K. 1910. Capsaicin, the pungent principle of *Capsicum*, and the detection of capsaicin. *Journal of Industrial and Engineering Chemistry* 2: 419–21.
- Oviedo y Valdés, G. F. de. [1557] 1950. *Sumario de la natural historia de las Indias*, ed. José Miranda. Mexico City.
- Pickersgill, B. 1984. Migrations of chili peppers, *Capsicum* spp., in the Americas. In *Pre-Columbian plant migration*, ed. Doris Stone, 106–23. Cambridge, Mass.
- Proctor, V. W. 1968. Long-distance dispersal of seeds by retention in digestive tract of birds. *Science* 160: 321–2.
- Purseglove, J. W. 1968. Some problems of the origin and distribution of tropical crops. *Genetics Agraria* 17: 105–22.
- Ridley, H. N. 1930. *The dispersal of plants through the world*. Ashford, England.
- Rozin, P. 1990. Getting to like the burn of chili pepper. In *Chemical senses*, ed. B. G. Green, J. R. Mason, and M. R. Kare, 231–69. New York.
- Ruskin, F. R., ed. 1990. *Lost crops of the Incas: Little-known plants of the Andes with promise for worldwide cultivation*. Washington, D.C.
- Ruiz, H., and J. Pavon. [1797] 1965. *Flora peruviana et chilensis*. 4 vols. Lehey, N.Y.
- Sauer, C. O. 1966. *The early Spanish Main*. Berkeley, Calif.
- Schurz, W. L. 1939. *The Manila galleon*. New York.
- Smith, P. G., and C. B. Heiser, Jr. 1951. Taxonomic and genetic studies on the cultivated peppers *C. annum* L. and *C. frutescens* L. *American Journal of Botany* 38: 367–8.
1957. Taxonomy of *Capsicum sinense* Jacq. and the geographic distribution of the cultivated *Capsicum* species. *Bulletin of the Torrey Botanical Club* 84: 413–20.
- Solanke, T. F. 1973. The effect of red pepper (*Capsicum frutescens*) on gastric acid secretion. *Journal of Surgical Research* 15: 385–90.
- Todd, P. H., Jr., M. C. Bensinger, and T. Biftu. 1977. Determination of pungency due to *Capsicum* by gas-liquid chromatography. *Journal of Food Science* 42: 660–5.
- Trease, G. E., and P. W. C. Evans. 1983. Drugs of biological origin. In *Pharmacognosy*. Twelfth edition, ed. G. E. Trease and P. W. C. Evans, 374–6. London.
- Turner, W. [1538] 1965. *Libellus de re herbaria*. London.
- Viranuvatti, V., C. Kalayasiri, O. Chearani, and U. Plengvanit. 1972. Effects of *Capsicum* solution on human gastric mucosa as observed gastroscopically. *American Journal of Clinical Nutrition* 5: 225–32.
- Watt, G. [1889] 1972. *A dictionary of the economic products of India*. Delhi.
- Watts, D. 1987. *The West Indies: Patterns of development, culture and environmental change since 1492*. Cambridge and New York.
- Willdenow, C. L. 1808. *Enumeratio plantarum horti regii botanici beroliensis*. 2 vols. Germany.

II.C.5 🌱 Cruciferous and Green Leafy Vegetables

Cruciferae (Brassicaceae), in the mustard family of the caper order (Capparales), are found on all continents except Antarctica. The cruciferae, so named because of the uniform, four-petaled flowers suggestive of a Greek cross, are an example of a natural family and demonstrate a large amount of diversity. Although most are weeds, the family includes significant food crop plants such as broccoli, cabbage, turnip, and radish. Cruciferae are most abundant in areas north of the equator and exhibit greatest variety in temperate and arid regions. The Mediterranean region is generally considered the site of the family's origination. Nonetheless, many of these cultigens appear to be native to northern Europe, and Reed C. Rollins (1993: 1) contends that the Irano-Turanian region of Eastern Europe and western Asia was the birthplace of at least some members of this plant family. A precise number of species and genera of the cruciferae is undetermined, although estimates range from 340 to 400 genera and 3,000 to 3,500 species (Vaughan, Macleod, and Jones 1976: vii; Rollins 1993: 2).

Taxonomy

The classification of cruciferae presents a challenge because of the large number of members and their unusually homogeneous nature (Hedge and Rechinger 1968: 1). But basic characteristics, both macroscopic and microscopic, mark the family as a whole. Typically the radial flower is characterized by the uniformity of its structure. This already mentioned flower type, four petals in the shape of a Greek cross,



Cabbage

is common to a large majority of this family's species. However, this pattern is altered by deviations, particularly in the structure of the stamen, flowers, and calyx. This is true of genera such as *Romanschulzia*, *Stanleya*, and *Warea*, and some species of *Streptanthus*, *Lepidium*, and *Megacarpaea* (Hedge and Rechinger 1968: 1-2; Rollins 1993: 2).

The fruits of the cruciferae, like the floral construction, are fundamentally homogeneous but can demonstrate significant variation in morphology. They play a key role in classification, along with developmental aspects of the plants, such as lifespan, floral maturation, seed germination, and possibly variations in sepal or petal formations (Hedge and Rechinger 1968: 1-2; Rollins 1993: 3).

In addition to a wide variety of macroscopic characteristics, several microscopic features may help identify the cruciferae group, such as the cell shape and configurations, as well as seed mucus (Hedge and Rechinger 1968: 2). A survey of the cruciferae group reveals one of the widest ranges of taxonomic characteristics among plant families, encompassing about 20 usable traits, sometimes existing in six or more states. Because of this high number of features, it is not unusual for authorities to emphasize different characteristics, resulting in, at times, a rather varied system of classification (Hedge and Rechinger 1968: 2).

The focus of this chapter is only on those genera and species associated with food or food products. Most of these species fall within *Brassica*, the best-known genus. Its 35 to 50 species and numerous varieties originated primarily in Europe, the Mediterranean, and Eurasia and include *Brassica oleracea* (cabbage, kale and collards, broccoli, cauliflower, Brussels sprouts, and kohlrabi, also known as *Brassica caulorapa*); *Brassica pekinensis* (Chinese cabbage); *Brassica nigra* (black mustard, also known as *Sinapis nigra*); *Brassica alba* (table mustard, also known as *Sinapis alba*); *Brassica juncea* (leaf mustard, also known as *Sinapis juncea*); *Brassica napobrassica* (rutabaga); and *Brassica rapa* or *Brassica*

campestris (turnips). There are, however, other significant food-producing members of this family, such as *Raphanus sativus* (radish) and *Nasturtium officinale* (watercress).

Cruciferae as a Cultivated Food Source

In Europe wild ancestors of the turnip and radish were gathered in prehistoric times, and most of these vegetables have been cultivated and used since the earliest days of recorded history. They are discussed extensively by classical Greek scholars like Theophrastus, and Roman writers, including Marcus Porcius Cato and Lucius Junius Moderatus Columella, as well as in chronicles of food and daily life in medieval and Renaissance Europe, such as *The Four Seasons of the House of Cerruti* (Spencer 1984). This book, compiled in the late 1300s, is based on the manuscript of an eleventh-century Arab physician living in northern Italy. The work describes the foods, drinks, and spices common in that region, along with practices considered good for health. Many cruciferous vegetables were grown during medieval and early modern times in the kitchen gardens of Europe, and particularly Britain, to be eaten in stews and salads.

Brussels sprouts



Then cabbage and its varieties were frequently referred to as "cole" or "coleworts," hence the name "coleslaw" for the popular side dish made with shredded cabbage. In Russia, cabbage, and to a lesser extent, turnips and radishes were important food crops. Along with radishes, a large number of *Brassica* varieties are found in China, where they have been used for centuries. Today, as in earlier periods, the durable and hardy plants of the Cruciferae family continue to play an important part in diets around the globe, and one that seems to increase in importance as more is learned of their nutritional and disease-preventive nature.

Cabbage and Its Varieties

Brassica oleracea includes some of the most significant vegetables used today, such as broccoli, cauliflower, Brussels sprouts, and of course, countless cabbages. With the exception of Brussels sprouts and kohlrabi, cabbage and its varieties have probably been cultivated since before recorded history (Toussaint-Samat 1992: 690). Wild cabbage, the early form of *B. oleracea*, was a small plant also known as "sea cabbage." Its leaves were firm and fleshy, fortified by mineral salts from the seawater it grew near, and even today *B. oleracea* can be found growing wild along the coasts of the English Channel.

Although there are approximately 400 species of cabbage, they can be divided into five groups. The first includes the familiar round, smooth-leaved cabbages that may be white, green, or red, as well as wrinkled-leaved varieties like Savoy. The second group comprises pointed cabbages like European spring and Chinese cabbages. A third category consists of cabbages with abnormally large, budding stems, as for example, Brussels sprouts. Green curly types such as kale represent a fourth group. These are used especially for animal food or for decoration of dishes for presentation, although kale is also featured in some famous soups, and collard greens make frequent appearances on many tables. The last category is made up of flowering cabbages such as cauliflower and broccoli (Toussaint-Samat 1992: 693).

Cabbage. Cabbage is the most durable and successful variety of *B. oleracea*. It is a versatile plant that can be found growing in almost every climate in the world, ranging from subarctic to semitropical. Such an ability to adapt to a wide variety of climatic conditions has enabled the vegetable to survive since prehistoric times. Although initially grown for its oily seeds, cabbage began to be used as a vegetable after people discovered that its green leaves were edible raw or cooked. Its consumption was confined to Asia and to Europe, however, as Neolithic Near Eastern peoples, Hebrews, and Egyptians did not use the plant.

In ancient Greece the writer Theophrastus noted three types of cabbage: curly-leaved, smooth-leaved, and wild. While comparing the curly-leaved and the

smooth-leaved varieties he observed that one bore either inferior seeds or none whatsoever. Unfortunately, he did not identify the one to which he referred, but he did say that the curly-leaved kind had better flavor and larger leaves than the smooth-leaved variety. Theophrastus described wild cabbage as having small round leaves with many branches and leaves. The plant had a strong medicinal taste and was used by physicians to ease or cure stomach problems (Theophrastus 1977, 2:85).

The Roman agronomist Cato the Elder (234–149 B.C.) also noted the medicinal value of cabbage, which, he contended, "surpasses [that of] all other vegetables." Whether eaten cooked or raw, cabbage was believed beneficial to digestion and to be an excellent laxative. Acknowledging the same three types of cabbage identified by Theophrastus, Cato agreed that the wild variety held the best medicinal value and wrote that it could be used as a poultice for all types of wounds, sores, or swellings. In addition, he advised "in case of deafness, macerate cabbage with wine, press out the juice, and instil warm into the ear, and you will soon know that your hearing is improved" (Cato 1954: 151, see also 141, 145). Both the Greeks and Romans believed that eating cabbage during a banquet would prevent drunkenness, and it has been pointed out that "the B vitamins contained in cabbage leaves do seem to have soothing and oxygenating qualities, very welcome when the mind is clouded by the fumes of alcohol. Research at a Texan [*sic*] university extracted a substance from cabbage which is useful in the treatment of alcoholism" (Toussaint-Samat 1992: 691).

In addition, cabbage was apparently inimical to fruits that could provide alcohol. Greek and Roman writers noted that cabbage was "hostile" to grapevines used for making wine. This is thought to be true even today: "Mediterranean farmers never plant it near vineyards in case bees transfer its odour to the bunches of grapes. Nor is it grown near beehives, because it might taint the flavour of the honey" (Toussaint-Samat 1992: 691).

Don Brothwell and Patricia Brothwell (1969) have written that the Romans favored two cabbage varieties known as *cymae* and *cauliculi* and pointed out that some scholars have mistaken *cauliculi* for Brussels sprouts when it was actually cabbage shoots or cabbage asparagus. *Cymae* is usually interpreted as sprouting broccoli and was apparently affordable only by the wealthier elements of Roman society. Moreover, by the time of Julius Caesar (100–44 B.C.), the Romans had enlarged cabbage, lavishing such attention on it and cultivating it to such a size that the poor of Rome could not afford to buy it (Toussaint-Samat 1992: 692). This interest in cabbage by the wealthy was apparently new because the vegetable had seldom been mentioned since the work of Cato, suggesting dietary distinctions between wealthier and poorer Romans that limited the consumption of ordi-

nary cabbage to the latter. According to Brothwell and Brothwell (1969: 118), "this is borne out by Juvenal's satire describing the differences between the food of the patron and that of his poor client - the patron has olives to garnish his excellent fish, the client finds cabbage in his 'nauseous dish.'"

Although the Romans introduced garden varieties of cabbage to northern Europe and Britain, it was not an entirely new food plant in these regions. On the basis of linguistic evidence, Anne C. Wilson (1974: 195) has pointed out that wild cabbage was used as a food source by Iron Age Celts living along the Atlantic coast of Europe prior to their migration to the British Isles. When the Romans did introduce their garden varieties of cabbage, she suggested the Celts favored an open-headed variety because of its similarity to this wild cabbage. However, due to the constant threat of famine during this era, the Celts continued to depend on the hardier wild variety as a safeguard against starvation (Wilson 1974: 196-7).

The fourteenth-century book *The Four Seasons*, mentioned previously, indicates that cabbage continued to enjoy a reputation for medicinal value in Renaissance Italy, although the work mentions some sources that thought cabbage bad for the blood and found its only redeeming quality to be its ability to "clear obstructions," of what kind we are left to wonder. On a less obscure note, cabbage was believed able to "restore a lost voice," and if its juice was cooked with honey and used sparingly as eyedrops it was believed to improve vision (Spencer 1984: 102).

Carroll L. Fenton and Herminie B. Kitchen (1956) divided cultivated cabbage into two main types - the hard-headed and the loose-headed, or Savoy cabbage. It is most likely that loose-headed cabbage evolved directly from wild cabbage found near the Mediterranean and Atlantic coasts of Europe. Its ridged leaves form a head at the top of a very short stalk. By contrast, hard-headed cabbage leaves are wound tightly around each other and around the stalk or "heart." It is believed that hard-headed cabbage developed in northern Europe. Because it does not grow well in warm climates, the Greeks and Romans did not cultivate it. Thus, it was probably developed by the Celts and has been cultivated by the Germans since ancient times. By the 1300s, hard-headed cabbage was common in England, and British soldiers introduced it in Scotland before 1650 (Fenton and Kitchen 1956: 74; Toussaint-Samat 1992: 692).

At first cabbage was an important crop in individual family plots known as kitchen gardens, but by the eighteenth century in England the cultivation of cabbage, along with many other vegetables, had expanded beyond kitchen gardens to the fields (Wilson 1974: 329; Braudel 1981, 1: 170). As early as 1540, Jacques Cartier grew hard-headed cabbages in Canada, and Native Americans used his seeds to plant

cabbages along with beans, squash, and corn (Fenton and Kitchen 1956: 74).

In the nineteenth and twentieth centuries the Russians have been among the world's most important consumers of hard-headed cabbage - an item of diet that has been a fundamental part of Russian cuisine for many centuries. It has been especially enjoyed pickled or prepared as cabbage soup called *shchii*. Usually flavored with meat fat or small chunks of meat, this soup consists of chopped cabbage, barley meal, salt, and a touch of *kvass* (Smith and Christian 1984: 252, 275-6).

Collards and kale. Collards (collard greens) and kale are varieties of *B. oleracea* that do not form heads. In fact, kale is very similar to sea cabbage (Fenton and Kitchen 1956: 72), and the primary difference between collard greens, a type of kale, and kale itself is leaf shape. Kale has a short, thick stalk and crinkly blue-green leaves that grow on leaf-stems, whereas collard greens have smooth, broad, yellowish green leaves (Fenton and Kitchen 1956: 72). Although the precise area of origin of collards and kale is unknown, it was most likely in Asia Minor or in the Mediterranean region, where both have been cultivated since prehistoric times. The Greeks and Romans grew several varieties of both kale and collard greens at least 2,200 years ago, followed about 200 years later by Germans and Saxons in northern Europe. They, or quite possibly the Romans, brought these plants to France and Great Britain. For nearly a thousand years kale and collards were the main winter vegetables in England. European colonists carried the seeds to the Americas. Kale and collards were cultivated in western Hispaniola before 1565, and by colonists in Virginia by at least 1669.

Most collards are similar in appearance, but kale has many varieties, some short and some very tall, such as a type of kale grown in England that reaches 8 or 9 feet in height. Today in the United States collards are grown predominantly in the South. An old and popular variety, 'Georgia collards', is characterized by stems 2 to 4 feet high, with leaves growing only at the top. Others include the 'Blue Stem', 'Green Glaze', 'Louisiana', and 'Vates Non-Heading'. Kale's principal types are 'Scotch', 'Blue', and 'Siberian' (Fenton and Kitchen 1956: 72-4; Carcione and Lucas 1972: 63-4).

Broccoli and cauliflower. Although well known today, broccoli and cauliflower are varieties of *B. oleracea* that are rarely mentioned in historical sources, despite being two of the oldest cultivated cabbage varieties. This may be because they were not well differentiated in those sources from the more recognizable cabbage. Jane O'Hara-May (1977: 251) has noted that in Elizabethan England the term "cabbage" referred to "the compact heart or head of the plant," whereas

the entire plant was known as cabbage-cole or colewort, a term applied to all varieties of cabbage.

Both broccoli and cauliflower, also called varieties of colewort, were cultivated "over 2,500 years ago in Italy or on the island of Cyprus" (Fenton and Kitchen 1956: 76), and broccoli, at least, was a part of Greek and Roman diets more than 2,000 years ago, although apparently it did not reach England until after 1700. When broccoli was introduced, in all likelihood it came from Italy because the English called it "Italian asparagus." In North America, broccoli was referred to in an 1806 book on gardening and grown regularly by Italian immigrants in private plots, but it was largely unknown to the public until the 1920s. It was because of effective marketing on the part of the D'Arrigo Brothers Company, which grew the vegetable, that demand for broccoli skyrocketed in the early 1930s, and it became "an established crop and an accepted part of the American diet" (Fenton and Kitchen 1956: 76; Carcione and Lucas 1972: 23). Today a major variety of broccoli is the 'Italian Green' or 'Calabrese', named after the Italian province of Calabria. Its large central head consists of bluish-green flower buds, called curds. 'De Cicco' is another popular variety that resembles the Calabrese but is lighter green in color. Chinese broccoli, also known as *Gai Lon*, "is more leaf than flower." It is light green in color, with small flower buds and large leaves on a long shank (Carcione and Lucas 1972: 23).

Through selective cultivation of sprouting broccoli, gardeners of long ago were able to produce ever larger clusters that were lighter in color and eventually became cauliflower broccoli and cauliflower. In ancient Greece cauliflower was popular, but after that its popularity declined in the West, where it was little used until the era of Louis XIV (Toussaint-Samat 1992: 691). The vegetable reached England toward the end of the Elizabethan era as part of an influx of new vegetables from Italy and France (Wilson 1974: 362; Braudel 1981, I: 223), and, in a list of 18 coleworts compiled in 1633, was identified as "Cole Florie" or "Colieflorie" and sold in London markets as "Cyprus coleworts" (O'Hara-May 1977: 251).

Although it is unclear when cauliflower arrived in North America, a number of varieties could be found in seed catalogs by the 1860s (Fenton and Kitchen 1956: 76; Carcione and Lucas 1972: 34). In all European countries, cauliflower is primarily a summer vegetable. It is very popular in Italy - the leading producer in Europe - where it comes in both bright purple and bright green colors. In the United States, typically, one can find only creamy, ivory-white cauliflower. Three of the most widely cultivated varieties in the United States are of the snowball type: 'Early Snowball', 'Super Snowball', and 'Snowdrift'. A larger, more leafy kind is the 'Danish Giant' grown primarily in the American Midwest (Carcione and Lucas 1972: 34; Toussaint-Samat 1992: 694).

Kohlrabi. A variety of *B. oleracea*, kohlrabi is just 400 to 500 years old and thus a relative newcomer to the cabbage genus of *Brassica*. It is one of the few vegetables with an origin in northern Europe. First described in 1554, kohlrabi was known in Germany, England, Italy, and Spain by the end of the sixteenth century. Documentation of its cultivation in the United States dates from 1806. Its common name, kohlrabi, is derived from the German *Kohl*, meaning cabbage, and *Rabi*, meaning turnip. It was developed by planting seeds from thick, juicy-stemmed cabbage, and the plants evolved into a turnip shape, characterized by slender roots at the bottom, a swelling of the stem into a turnip-sized bulb just above the ground, and leaves similar to those of a turnip sprouting on top (Fenton and Kitchen 1956: 77; Carcione and Lucas 1972: 66). Although more delicate, the kohlrabi's taste resembles that of a turnip. Europeans grow frilly-leaved varieties of kohlrabi for ornament, whereas in the United States the two common varieties are both food plants (Carcione and Lucas 1972: 66).

Brussels sprouts. Less than 500 years old and native to northern Europe, Brussels sprouts are also a recent addition to *B. oleracea*. Described as early as 1587, the plant supposedly got its name as the result of its development near the city of Brussels, Belgium. Brussels sprouts were cultivated in England in the seventeenth century and appear to have been introduced into the United States in the nineteenth century, although exactly when, where, and by whom is unclear (Fenton and Kitchen 1956: 76-7; Carcione and Lucas 1972: 25; Wilson 1974: 203).

Mustard

Often found growing in fields and pastures, the mustard varieties *B. nigra* and *B. alba* are characterized by leaves with deep notches and small yellow flowers with four petals forming the shape of a Greek cross, typical of the Cruciferae (Fenton and Kitchen 1956: 66). These mustard varieties evolved from weeds growing wild in central Asia into a food source after humans learned that their pungent seeds improved the taste of meat - a not unimportant discovery in ancient times, when there was no refrigeration and meat was usually a bit tainted (Fenton and Kitchen 1956: 66). Once mustard became recognized as a spice, it was commercially cultivated, and traders carried the seed to China and Japan, Africa, Asia Minor, and Europe.

The Greeks and the Romans first used mustard for medicinal purposes by creating ointments from the crushed seeds and prescribing the leaves as a cure for sore muscles. According to Pliny the Elder, the first-century Roman writer, mustard "cured epilepsy, lethargy, and all deep-seated pains in any part of the body" (1938, I: 64), and also, mustard was an "effective cure for hysterical females" (Carcione and Lucas 1972: 64). In addition to medical applications, the Greeks

and Romans came to use the seeds as a spice and the boiled leaves as a vegetable. The Romans also pulverized mustard seeds and added them to grape juice to prevent it from spoiling. This practice later appeared in England, where grape juice was called "must" and both seeds and plants were known as mustseed. Over time, the spelling evolved from "mustseed" to "mustard" (Fenton and Kitchen 1956: 66–7).

By the time of the Middle Ages, mustard was popular as a condiment that seasoned food and stimulated the appetite; it was also used to treat gout and sciatica and as a blood thinner. Caution was advised, however, in smelling mustard powder, although the risk of its rising to the brain could be averted by using almonds and vinegar in its preparation (Spencer 1984: 47).

Of the two varieties of mustard that were cultivated in the United States before 1806, both "ran wild," becoming weeds, as did another type, the Indian mustard, grown for greens (Fenton and Kitchen 1956: 67–8). Today, India, California, and Europe supply most of the world's mustard. Joe Carcione and Bob Lucas noted that wild mustard, also called "Calutzi," colors the California hills a brilliant yellow in springtime. Commercially grown types include 'Elephant Ears', which have large plain leaves, and the curly-leaved varieties, 'Fordhook Fancy' and 'Southern Curled' (Carcione and Lucas 1972: 64). The dry seeds are crushed to produce oil and ground into powder, which is the basis of the condiment.

Chinese Brassica

A large variety of green vegetables are grown in China, and prominent, particularly in the north, are several types of *Brassica* that are native Chinese cultigens. As with cruciferous vegetables in general, their exact classifications are controversial and difficult to sort out. The most common are *B. pekinensis*, *Brassica chinensis*, and *B. juncea*. The long, cylindrical-headed *B. pekinensis*, known as *Pai ts'ai*, or Chinese cabbage, and by numerous colloquial names in different languages, has white, green-edged leaves wrapped around each other in a tall head reminiscent of celery. The nonheaded *B. chinensis*, identified as *ch'ing ts'ai* in Mandarin and *pak choi* in Cantonese, has dark green, oblong or oval leaves resembling chard, growing on white stalks. Descriptions of *pak choi* and *pai ts'ai* exist in Chinese books written before the year A.D. 500, although both were probably developed even earlier. However, it was not until the early twentieth century that they became commonly known in North America (Fenton and Kitchen 1956: 68; Anderson and Anderson 1977: 327), ironically at a time when Chinese immigration into the United States and Canada was all but nonexistent. It is possible that their expanded use in North America came with the growth of second and third generations of Chinese populations in North America.

Brassica juncea, recognized as *chieb ts'ai* in Mandarin and *kaai choi* in Cantonese, has characteristics

similar to *B. chinensis*. Also important in China are *Brassica alboglabra*, named *kaai laan* in Cantonese, which are similar to collard greens; *B. campestris*, the Chinese rapeseed that is a major oil-producing crop (canola oil in the West); and several minor crops including *B. oleracea*, recently introduced from the West. Although quite distinctive, the Cantonese *choi sam* or "vegetable heart" is considered to be a form of *B. chinensis* (Anderson and Anderson 1977: 327).

Along with radishes, these *Brassica* cultigens are the most significant "minor" crops grown in southern China and, combined with rice and soybeans, constitute the diet of millions. Cabbage or mustard greens stir-fried in canola oil and seasoned with chillies or preserved soybean constitutes a nutritionally sound meal without the use of animal products or plants that require large areas of land and are overly labor intensive. Chinese *Brassica* varieties produce large yields and are available throughout the year, particularly in south China (Anderson and Anderson 1977: 328).

Radish

According to Reay Tannahill (1988: 11), radishes, *R. sativus*, were part of the diet of prehistoric hunter-gatherer cultures of Europe and have been grown and eaten, especially pickled, in the Orient for thousands of years. Because of the radish's antiquity and the many varieties that have been cultivated all over the Old World, including the Orient, its precise origin is obscure. Early radishes were probably large enough to be used as a food and not merely as a salad decoration or appetizer. The leaves may also have been eaten as greens (Brothwell and Brothwell 1969: 110).

Cultivated radish varieties were transported across Asia to Egypt about 4,000 years ago, where the Egyptians "ate radish roots as vegetables and made oil from the seeds" (Fenton and Kitchen 1956: 69; see also Darby, Ghalioungui, and Grivetti 1977, II: 664). Two radishes were discovered in the necropolis of Illahun (Twelfth Dynasty), and the leaves and roots of the specific Egyptian radish variety *aegyptiacus* have been identified (Darby et al. 1977: 664). Fenton and Kitchen claimed that pictures of radishes were chiseled into the walls of a temple at Karnak, on the River Nile (Fenton and Kitchen 1956: 69). According to William J. Darby, Paul Ghalioungui, and Louis Grivetti, however, the evidence of radish use in ancient Egypt is primarily literary, particularly from Pliny's *Natural History*. They pointed out that in all likelihood radishes in Egypt were valued for oil produced from their seeds rather than as a food product. The Egyptians also considered the radish to be of medicinal value in curing a now unknown disease called *Phtheiriasis*. Poorer Egyptians employed radish oil as an inexpensive method of embalming, using it as an enema for emptying the intestines (Darby et al. 1977: 664, 785).

Theophrastus observed that in ancient Greece there existed five varieties of radishes: 'Corinthian', 'Cleonae', 'Leiothasian', 'Amorea', and 'Boeotian'. He noted that those types with smooth leaves had a sweeter and more pleasant taste, and those having rough leaves tasted sharp. Unfortunately, he did not associate varieties with leaf type (Theophrastus 1977, 2: 81-3). Fenton and Kitchen suggested that the Greeks were so fond of radishes that they used golden dishes to offer them to their god Apollo, whereas silver dishes were sufficient for beets, and turnips warranted only bowls of lead (Fenton and Kitchen 1956: 69).

Columella's instructions about the cultivation of radishes establishes their presence in Rome (Columella 1960, 3: 157, 165-9), and Pliny also wrote about the common use of radishes. He indicated that they were grown extensively for their seed oil, but that as a food, he found them a "vulgar article of diet" that "have a remarkable power of causing flatulence and eructation" (Brothwell and Brothwell 1969: 110).

Radishes were introduced to England by the occupying Romans and were known to Anglo-Saxon and Germanic peoples by the early medieval era. Like cabbage, radishes were common in English kitchen gardens by the fifteenth century (Wilson 1974: 196-7, 205), and they reached England's American colonies early in the seventeenth century. Europeans occasionally ate raw radishes with bread, but more common was the use of the roots in a sauce served with meat to stimulate the appetite. So highly did the great Italian composer Gioacchino Rossini regard radishes that they were one of the *Four Hors d'Oeuvres* in his famous opera. For others, however, radishes have inspired mistrust. A *Plague Pamphlet*, printed in London in 1665, noted that the appearance of the dread disease was the result of "eating radishes, a cat catter wouling, . . . immoderate eating of caviare and anchoves, tame pigeons that flew up and down an alley, [and] drinking strong heady beer" (Carcione and Lucas 1972: 104).

Radishes are used today mostly as an appetizer and in salads, and sometimes the young, tender leaves are boiled and served as greens. Radish varieties are numerous, and they come in many sizes (ranging from cherry- to basketball-size), shapes (round, oval, or oblong), and colors (white, red and white, solid red, or black). But their flavors are very similar. Most common in the United States are small, round, red or white varieties, including 'Cherry Belle' and 'Comet'. The 'White Icicle' is a long and narrow white radish. Oriental radishes are the most spectacular, with some, like the Japanese daikon, reaching several feet in length (Fenton and Kitchen 1956: 69; Carcione and Lucas 1972: 104). Europeans grow large, hot-tasting winter radishes, which they store in cool, dark cellars and eat during cold weather. According to an old description, winter radishes could reach a weight of 100 pounds (Fenton and Kitchen 1956: 69). Methods

of serving radishes in the various parts of Asia include pickling them in brine, boiling them like potatoes, eating them raw, and cooking them as fresh vegetables (Fenton and Kitchen 1956: 69-70).

Turnip

Grown since ancient times, the turnip, *B. rapa* or *B. campestris*, has long been prized as a staple winter food, and in some areas it has been the only winter produce available. According to Brothwell and Brothwell, turnip varieties seem to have been indigenous to the region between the Baltic Sea and the Caucasus, later spreading to Europe (Brothwell and Brothwell 1969: 110). Today turnips continue to grow wild in eastern Europe and Siberia. They are almost perfectly round and have white flesh and thin, rough leaves covered by prickly hairs (Fenton and Kitchen 1956: 70). Their cultivation predates recorded history, and excellent storing qualities must have made the vegetable a dependable winter food for livestock as well as people (Brothwell and Brothwell 1969: 110-11).

In antiquity, the name "turnip" also referred to radishes and other root vegetables save leeks and onions (Darby et al. 1977, II: 665). Several varieties of turnips - round, long, and flat - were used by the Romans prior to the Christian Era. Greek and Roman writers indicated that the use was limited largely to "the poorer classes and country folk." Theophrastus wrote that there was disagreement over the number of varieties; he also provided some instructions for their cultivation. He stated that like the radish, the turnip's root grew best and sweetest in wintertime (Theophrastus 1977, 2: 83). Columella pointed out that turnips should not be overlooked as an important crop because they were a filling food for country people and a valuable source of fodder for livestock. In addition, Columella provided his readers with a recipe for pickling turnips in a mustard and vinegar liquid (Columella 1960, 1: 171, 3: 331-3). Apicius recommended mixing them with myrtle berries in vinegar and honey as a preservative. Pliny considered them the third most important agricultural product north of the Po River, and wrote that the leaves were also eaten. Especially interesting was his opinion that turnip tops were even better tasting when they were yellow and half dead (Pliny 1938, 5: 269-71; Brothwell and Brothwell 1969: 111).

The medieval chronicle *The Four Seasons* noted that if soaked in vinegar or brine, turnips could be preserved for up to a year. Sweet-tasting and thin-skinned types were considered the best. Medicinally, the turnip was believed good for the stomach, capable of relieving constipation, and effective as a diuretic. In preparing turnips, the advice was for prolonged cooking, even cooking them twice, to avoid indigestion, flatulence, and swelling. If these problems did occur, however, an emetic of vinegar and salt was recommended as a remedy (Spencer 1984: 109).

Like cabbage and radishes, turnips were a part of

vegetable gardens in Roman Britain. By A.D. 1400, they were common in France, Holland, and Belgium, and at that date were among a quite small number of vegetables that had been known and available in northern Europe for centuries (Fenton and Kitchen 1956: 70; Carcione and Lucas 1972: 123). Explorers and colonists brought turnips to North America in the late sixteenth and early seventeenth centuries, where, because they flourish in cool weather, they became a summer crop in the north and a winter crop in the south. Modern varieties are generally less than 5 inches thick, but a turnip weighing 100 pounds was once grown in California, and during the 1500s, most European turnips weighed 30 to 40 pounds (Fenton and Kitchen 1956: 71). Commercially, there are many varieties grown today, and although shape and skin color may differ, like radishes the taste remains the same. The 'Purple-Top White Globe' and the 'Purple-Top Milan' are grown for their roots, and 'Shogoin', an Oriental variety, is harvested for its tender greens (Carcione and Lucas 1972: 123-4).

Rutabaga

Rutabagas are occasionally referred to as "Swede turnips" or just "Swedes" because they were developed and grown in Sweden before A.D. 1400. According to Carcione and Lucas (1972: 123), rutabagas appear to be "the result of a meeting of a swinging Swedish turnip and an equally willing cabbage." Although closely related to the turnip, the rutabaga, *B. napobrassica*, is a relatively modern vegetable that is larger and longer than the turnip, with a milder taste and flesh that is yellow-colored rather than white. In addition, its leaves are smooth and thick, whereas turnip leaves are thin, rough, and prickly (Fenton and Kitchen 1956: 70). Cattle and pigs feed on raw rutabagas, but people eat the roots boiled and buttered (Fenton and Kitchen 1956: 71; Drummond and Wilbraham 1991: 180).

Rutabagas spread from Sweden to central Europe and the northern regions of Italy, where in medieval times they were called "Swedes" and housewives were advised to accept only those that were garden-fresh. Although "Swedes" reputedly bloated the stomach, they were delicious when prepared with meat broth. Moreover, they were thought to "activate the bladder," and "if eaten with herbs and abundant pepper, they arouse young men to heights of sexual adventurousness" (Spencer 1984: 58).

Rutabagas were cultivated prior to 1650 in Bohemia, and in 1755 they were introduced to England and Scotland from Holland, where they were initially referred to as "turnip-rooted cabbage," "Swedes," or "Swedish turnips." American gardeners were growing rutabagas by 1806, and today, Canada - along with the states of Washington and Oregon - supplies the United States with rutabagas, which explains their often being called "Canadian turnips" by Americans. Two of the best-known rutabaga types are the

'Laurentian' and the 'Purple-Top Yellow' (Carcione and Lucas 1972: 123-4).

Watercress

Native to Asia Minor and the Mediterranean region, watercress (*Nasturtium officinale*) grows wild wherever shallow moving water is found. It is characterized by long stems and small thick leaves (Carcione and Lucas 1972: 126). According to Brothwell and Brothwell, the Romans consumed watercress with vinegar to help cure unspecified mental problems, and both Xenophon, the ancient Greek general, and the Persian King Xerxes required their soldiers to eat the plant in order to maintain their health (Brothwell and Brothwell 1969: 122; Carcione and Lucas 1972: 126). Ancient cress seeds found in Egypt probably arrived from Greece and Syria, where cress is found among a list of Assyrian plants. Dioscorides maintained that watercress came from Babylon (Brothwell and Brothwell 1969: 122-3).

Stronger-flavored kinds of watercress were preferred in medieval Italy, where they allegedly provided a variety of medicinal benefits. Although watercress was blamed for headaches, it supposedly strengthened the blood, aroused desire, cured children's coughs, whitened scars, and lightened freckles. Additionally, three leaves "picked with the left hand and eaten immediately will cure an overflow of bile" (Spencer 1984: 19).

Cultivation of watercress for sale in markets dates to about 1800 in England, although wild watercress was doubtless gathered by humans for many millennia. Brought to the United States by early settlers, watercress can now be found throughout the country. Soil-cultivated relatives of watercress include peppergrass (also called curly cress), upland cress, lamb's cress, cuckoo flower, lady's smock, mayflower, pennycress, and nasturtiums (Carcione and Lucas 1972: 126).

Nutrition, Disease Prevention, and Cruciferous Vegetables

Cruciferous vegetables have substantial nutritional value. They contain significant amounts of beta-carotene (the precursor of vitamin A), vitamin C (ascorbic acid), and "nonnutritive chemicals" such as indoles, flavones, and isothiocyanates, which contribute to the prevention of diet-related diseases and disorders such as blindness and scurvy. In addition, recent investigations have shown them to be effective in warding off several types of cancer. Studies have linked diets high in vitamin A to cancer prevention, and research also indicates that chemicals like indoles inhibit the effects of carcinogens. Vitamin C is recognized as an effective antioxidant, which is thought to be preventive against the development of some cancers and an inhibitor of the progress of the human immunodeficiency virus (HIV).

Moreover, the substances known as antioxidants are critical in the maintenance of homeostasis, a state of physiological equilibrium among the body's functions and its chemical components. Molecules that form the human body are typically held together by the magnetic attraction between their electrons. Occasionally, however, these molecules exist in an oxidized state, meaning that they have unpaired electrons and are seeking out "molecular partners," often with potentially harmful effects. In this condition, these molecules are known as "free radicals," and because they can react more freely with their surrounding environment, they are capable of disrupting many finely tuned processes essential in maintaining good health. In some cases, free radicals serve no particular function and are simply the waste products of bodily processes, but in others, the body's immune system uses them to fight diseases. Yet even when useful, free radicals can damage nearby tissue and impair bodily functions. To control such damage, the human body uses antioxidants to neutralize the effects of free radicals. Unfortunately, there are often inadequate amounts of antioxidants to eliminate all of the free radicals, and there are also periods or conditions during which the number of free radicals increases along with the damage they inflict. This is particularly true when people are infected with HIV or have developed cancer (Romeyn 1995: 42-3).

In the cases of both HIV and cancer, vitamin C reacts with and neutralizes free radicals. This vitamin also helps increase the overall antioxidant ability of vitamin E by preventing certain functions of vitamin E that can actually inhibit the effects of its antioxidant characteristics. Specifically with regard to cancer prevention, vitamin C, of which cruciferous vegetables have high levels, appears to significantly reduce the risk of contracting stomach or esophageal cancers. Other, less conclusive studies suggest that vitamin C may also inhibit the development of bladder and colon cancer. In addition to its role as an antioxidant, vitamin C acts to inhibit the formation of cancer-causing nitrosamines, which are created by cooking or by digesting nitrites found in food.

In citing over a dozen studies worldwide, Patricia Hausman (1983: 24-5) noted that diets rich in vitamin A provide a surprising amount of protection from cancer in eight different organs. The strongest evidence links vitamin A to the prevention of lung, stomach, and esophageal cancer. Although less conclusive, other studies have recognized the potential of vitamin A to protect against cancer of the mouth, colon, rectum, prostate, and bladder.

The term, "vitamin A", encompasses many substances that can fulfill the body's requirements for this nutrient. Retinol is the form found in foods derived from animal products. Beta-carotene and carotenoids are found in fruits and vegetables; however, carotenoids are only a minor source. For most

bodily functions that require vitamin A, any one of these forms will suffice, but it is beta-carotene that is tied most closely to cancer prevention (Hausman 1983: 24-5).

Although it is unclear how beta-carotene aids in the prevention of cancer, some chemists have suggested that it might act as an antioxidant. However, Eileen Jennings has noted that there is little doubt concerning the significance of vitamin A as it relates to gene regulation: "The gene regulator and antiproliferation effects of vitamin A may be the entire explanation for the anticancer effect of vitamin A" (1993: 149). Although the final determination of beta-carotene's antioxidant qualities awaits further study, its ability to act as a cancer preventive has been demonstrated, and thus, it is recommended that the cruciferous vegetables containing high levels of this substance be eaten frequently.

Scientific studies also have demonstrated that cruciferous vegetables further limit cancerous growth because they contain small quantities of indoles, flavones, and isothiocyanates. According to Jennings, these nonnutritive chemicals have been shown "to either stimulate production of enzymes that convert toxic chemicals to less toxic forms or interfere with the reaction of carcinogens with DNA" (1993: 223). Hausman (1983: 82-3) wrote that the enzyme system that produces these cancer inhibitors is recognized as the "mixed function oxidase system." Because the family of cruciferous vegetables contains such high levels of these inhibitors, particularly indoles as well as high levels of beta-carotene and vitamin C, the Committee on Diet, Nutrition, and Cancer of the National Academy of Sciences in 1982 emphasized eating these vegetables often. Broccoli, cauliflower, Brussels sprouts, and cabbage have all been linked to lowering the risk of developing stomach and colon cancer, and some studies indicated a possible connection to a reduced risk of rectal cancer.

The absence of beta-carotene and vitamin C in the diet is also linked to the development of deficiency diseases such as "night blindness" and scurvy. Because of the unavailability of fruits or cruciferous vegetables and other foods containing these nutrients, in many parts of the developing world deficiency diseases remain common. Extended vitamin A deficiency results in a severe defect called xerophthalmia, which affects the cornea of the eye. More often this deficiency causes "night blindness," or nyctalopia, which is the inability to see in a dim light. This problem has been very common in the East for several centuries, and according to Magnus Pyke (1970: 104, 127), at least until the past two or three decades, vitamin A deficiency annually caused thousands of cases of blindness in India.

The earliest prescribed treatment of an eye disorder thought to be night blindness is found in the Ebers Papyrus, an Egyptian medical treatise dating from around 1600 B.C. Rather than prescribing the

regular consumption of cruciferous vegetables, however, it suggested using ox liver, itself high in vitamin A, which it claimed was “very effective and quick-acting.” Hippocrates and Galen were also familiar with night blindness, the former recommending the consumption of ox liver (in honey) as a remedy. Later Roman medical writings gave similar advice, and Chinese literature contained descriptions of this eye condition by A.D. 610 (Brothwell and Brothwell 1969: 179–80). As already noted, the work on foods in medieval Italy, *The Four Seasons*, contended that the cooked juice of cabbage mixed with honey and used sparingly as an eyedrop improved vision (Spencer 1984: 102).

Like night blindness, scurvy, a disease caused by an insufficient intake of vitamin C, continues to be a serious deficiency disease of the developing world. One of the primary functions of vitamin C is the production of collagen, which helps to build new tissue, in a sense maintaining the very structure of the human body. It can reduce lung tissue damage caused by the activation of the body's immune system and is important to the production of hormones, steroids, and neurotransmitters. It is also required for the “conversion of folate into its active form” and aids in iron absorption (Pyke 1970: 114; Hausman 1983: 43; Romeyn 1995: 44–5). A prolonged lack of vitamin C causes the weakening and breakdown of the body's cell structure and tissue, with scurvy the visible symptom of this phenomenon.

Interestingly, humans are susceptible to scurvy only because of an unfortunate biochemical shortcoming in their genetic makeup. Unlike most animals, humans and four other species are unable to synthesize vitamin C and therefore will suffer from scurvy if the vitamin C intake from their food is insufficient for a long enough period of time (Pyke 1970: 115).

Brothwell and Brothwell suggest that deficiencies in vitamin C were rare prior to the appearance of urban development (beginning in the Neolithic era) because hunter-gatherers had more diversity in their diet. Hippocrates has been credited with the earliest mention of scurvy when he described an unpleasant condition characterized by “frequent haemorrhages” and “repulsive ulceration of the gums” – both features of the disease. Pliny also acknowledged the presence of this condition (which he called *stomacace*) in Roman troops stationed in the Rhine region. Writings from the Middle Ages contain many references to scurvy as well, implying that it was prevalent in the Europe of that era (Brothwell and Brothwell 1969: 181).

Before the introduction of the white potato, the disease was common in the spring in the northern countries of Europe. It has also ravaged sailors during long sea voyages, as well as arctic explorers whose provisions consisted mostly of easily preserved food and lacked fresh fruits and vegetables (Pyke 1970: 113–14). But although the disease is of

“considerable antiquity,” scurvy occurs infrequently today despite the large numbers of underfed and malnourished people in the world because it is a disease that requires an unusual state of deprivation (Pyke 1970: 112). However, scurvy in children has been reported in recent years in Toronto, Canada, in many communities in India, and in Glasgow, Scotland, and “bachelor scurvy” can occur in instances where older men who live alone fail to consume meals with adequate quantities of vegetables (Pyke 1970: 115).

In order to improve nutritional levels in humans and thus prevent deficiency diseases, as well as improve overall health, Henry M. Munger (1988) has suggested that instead of breeding crops to increase nutrient levels, a more effective, and less expensive, solution is to increase consumption of plants already high in nutritional value. He offered broccoli as an example. Relatively unknown in the United States 50 years ago, broccoli has experienced a dramatic increase in its production, from approximately 200 million pounds in the 1950s to over a billion pounds in 1985. In a 1974 study by M. Allen Stevens on *Nutritional Qualities of Fresh Fruits and Vegetables* that compares the nutritional values of common fruits and vegetables, broccoli was ranked highest in nutritional value because of its substantial content of vitamins A and C, niacin, riboflavin, and nearly every mineral. Additionally, “based on dry weight, broccoli contains protein levels similar to soybean” (Stevens 1974: 89). In an earlier study by Stevens, however, broccoli was ranked twenty-first in contribution to nutrition, based on a formula derived from its production as well as its nutrient content. But by 1981, that ranking had risen to seventh, a direct result of its increased production and consumption levels (Stevens 1974: 89–90; Munger 1988: 179–80).

Munger has maintained that improved nutrition in tropical developing countries can be achieved by adapting nutritious, temperate crops to tropical conditions or by expanding the use of less familiar, highly nutritious tropical vegetables. Two of his primary illustrations involve cruciferous vegetables. In the first case he cited an adaptation of Japanese cabbage hybrids for cultivation in the more tropical lowlands of the Philippines. This move lowered production costs and made cabbage more accessible and affordable to the general population. A second example, *choi-sum* or Flowering White Cabbage, a variety of *B. campestris*, is similar to broccoli and grown near Canton, China. Nearly all parts of this highly nutritious plant are edible. It can be planted and harvested year-round, and its production yield is comparable to that of potatoes and corn in the United States. Because of its efficient production and high nutrient concentration, *choi-sum* also seems to Munger a good candidate for promotion in tropical developing nations (Munger 1988: 180–1, 183).

Summary

The cruciferous family of vegetables includes some of the most nutritionally significant foods produced today. Predominantly European and Asian in origin, these vegetables have a history of cultivation and use that spans many centuries. The ancient Greeks and Romans employed some of them not only as food-stuffs but also for medicinal purposes. They believed that cabbage, for example, could cure a wide range of ailments, from healing wounds to correcting problems with internal organs. In medieval and Renaissance Europe as well as in Russia and China, cruciferous vegetables were found in kitchen gardens and composed an important part of the daily diet. Gradually they were transformed from garden produce into commercial crops and today are abundantly available for sustenance and for good health. Contemporary research suggests a link between cruciferous vegetables and disease prevention. Because of their high levels of vitamin C, beta-carotene, and other disease inhibitors, these food plants help avoid deficiency diseases, prevent some cancers, and retard the development of HIV in the human body. Such findings suggest that the consumption of cruciferous vegetables has a positive effect on health, and consequently they should have a prominent place in the human diet.

Robert C. Field

Bibliography

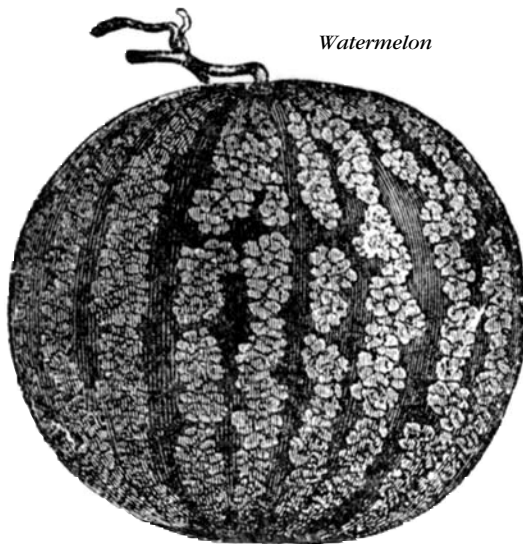
- Anderson, E. N., Jr., and Marja L. Anderson. 1977. Modern China: South. In *Food in Chinese culture: Anthropological and historical perspectives*, ed. K. C. Chang, 319–82. New Haven, Conn.
- Baldinger, Kathleen O'Bannon. 1994. *The world's oldest health plan*. Lancaster, Pa.
- Braudel, Fernand. 1981. *Civilization and capitalism, 15th–18th century*, trans. Sian Reynolds. 3 vols. New York.
- Brothwell, Don, and Patricia Brothwell. 1969. *Food in antiquity: A survey of the diet of early peoples*. New York.
- Carcione, Joe, and Bob Lucas. 1972. *The greengrocer: The consumer's guide to fruits and vegetables*. San Francisco, Calif.
- Cato, Marcus Porcius. 1954. *On agriculture*, trans. William Davis Hooper, rev. by Harrison Boyd Ash. London.
- Columella, Lucius Junius Moderatus. 1960. *On agriculture and trees*, trans. Harrison Boyd Ash, E. S. Forster, and Edward H. Heffner. 3 vols. London.
- Darby, William J., Paul Ghalioungui, and Louis Grivetti. 1977. *Food: The gift of Osiris*. 2 vols. London.
- Drummond, J. C., and Anne Wilbraham. 1991. *The Englishman's food: A history of five centuries of English diet*. London.
- Fenton, Carroll Lane, and Herminie B. Kitchen. 1956. *Plants that feed us: The story of grains and vegetables*. New York.
- Hausman, Patricia. 1983. *Foods that fight cancer*. New York.
- Hedge, Ian C., and K. H. Rechinger. 1968. *Cruciferae*. Graz.
- Jennings, Eileen. 1993. *Apricots and oncogenes: On vegetables and cancer prevention*. Cleveland, Ohio.
- McLaren, Donald S., and Michael M. Meguid. 1988. *Nutrition and its disorders*. Fourth edition. Edinburgh.
- Munger, Henry M. 1988. Adaptation and breeding of vegetable crops for improved human nutrition. In *Horticulture and human health: Contributions of fruits and vegetables*, ed. Bruno Quebodeaux and Fredrick A. Bliss, 177–84. Englewood Cliffs, N.J.
- O'Hara-May, Jane. 1977. *Elizabethan dyetary of health*. Lawrence, Kans.
- Pliny the Elder. 1938. *Pliny: Natural history*, trans. H. Rackham. 10 vols. Cambridge, Mass.
- Pyke, Magnus. 1970. *Man and food*. New York.
- Rollins, Reed C. 1993. *The cruciferae of continental North America: Systematics of the mustard family from the Arctic to Panama*. Stanford, Calif.
- Romeyn, Mary. 1995. *Nutrition and HIV: A new model for treatment*. San Francisco, Calif.
- Smith, R. E. F., and David Christian. 1984. *Bread and salt: A social and economic history of food and drink in Russia*. Cambridge and New York.
- Spencer, Judith, trans. 1984. *The four seasons of the house of Cerruti*. New York.
- Stevens, M. Allen. 1974. Varietal influence on nutritional value. In *Nutritional qualities of fresh fruits and vegetables*, ed. Philip L. White and Nancy Selvey, 87–110. Mount Kisco, N.Y.
- Tannahill, Reay. 1988. *Food in history*. New York.
- Theophrastus. 1977. *Enquiry into plants*, 2 vols., trans. Sir Arthur Hort. Cambridge, Mass.
- Toussaint-Samat, Maguelonne. 1992. *A history of food*, trans. Anthea Bell. Cambridge, Mass.
- Vaughan, J. G., A. J. Macleod, and B. M. G. Jones, eds. 1976. *The biology and chemistry of the Cruciferae*. London and New York.
- Wilson, C. Anne. 1974. *Food and drink in Britain from the Stone Age to recent times*. New York.
- Zohary, Daniel, and Maria Hopf. 1993. *Domestication of plants in the Old World*. Oxford.

II.C.6 🍆 Cucumbers, Melons, and Watermelons

Our focus here is on three important cucurbits – cucumber, melon, and watermelon – although cucurbits of less significance such as the citron, bur (or West India gherkin), and some lesser-known melons are also briefly discussed. These plants, together with all the sundry squashes and pumpkins, constitute a taxonomic group of diverse origin and genetic composition with considerable impact on human nutrition. The term “cucurbit” denotes all species within the Cucurbitaceae family.



Cucumber



Cucurbits are found throughout the tropics and subtropics of Africa, southeastern Asia, and the Americas. Some are adapted to humid conditions and others are found in arid areas. Most are frost-intolerant so they are grown with protection in temperate areas or to coincide with the warm portion of the annual cycle. Cucurbits are mostly annual, herbaceous, tendril-bearing vines.

The significance of cucurbits in human affairs is illustrated by the abundance of literature devoted to them, albeit much less than that produced on the

grains and pulses. Two full-length books have cucurbits as the title (Whitaker and Davis 1962; Robinson and Decker-Walters 1997), and at least four significant publications have been derived from recent conferences on these plants (Thomas 1989; Bates, Robinson and Jeffrey 1990; Lester and Dunlap 1994; Gómez-Guillamón et al. 1996). Moreover, a recent reference book provides an inclusive chapter on cucurbits (Rubatzky and Yamaguchi 1997) and an annual publication is dedicated to their genetics (Ng 1996).

Taxonomy

The Cucurbitaceae family is well defined but taxonomically isolated from other plant families. Two subfamilies - Zanonioideae and Cucurbitoideae - are well characterized: the former by small, striate pollen grains and the latter by having the styles united into a single column. The food plants all fall within the subfamily Cucurbitoideae. Further definition finds cucumber (*Cucumis sativus* L.) and melon (*Cucumis melo* L.) to be within the subtribe Cucumerinae, tribe Melothrieae, and watermelon (*Citrullus lanatus* [Thunb.] Matsum. and Nakai.) is assigned to the tribe Benincaseae, subtribe Benincasinae. The taxonomic sites of West India gherkin (*Cucumis anguria* L.) and citron (*Citrullus lanatus* var. *citroides* [L.H. Bailey] Mansf.) are, as with those just listed, in the same genus. There are about 118 genera and over 800 species in the Cucurbitaceae (Jeffrey 1990a). The melons (*C. melo*) are further subdivided into groups that do not have taxonomic standing but have proved useful horticulturally (Munger and Robinson 1991):

Figure II.C.6.1. Netted cantaloupe fruit.



The Cantalupensis group includes cantaloupe, muskmelon (Figure II.C.6.1), and Persian melon. The fruit are oval or round; sutured or smooth; mostly netted, some slightly netted or nonnetted; and abscise from the peduncle when mature. The flesh is usually salmon or orange colored, but may be green and is aromatic. In the United States, the term “muskmelon” and “cantaloupe” may be used interchangeably, but some horticultural scientists (Maynard and Elmstrom 1991: 229) suggest that they be used to distinguish between types of *C. melo* Cantalupensis group. This group includes the previously recognized Reticulatus group.

The Inodorus group consists of winter melon, casaba (Figure II.C.6.2), crenshaw, honeydew, Juan Canary (Figure II.C.6.3), and Santa Claus (Figure II.C.6.4). The fruit are round or irregular, smooth or wrinkled, but not netted; nor do they abscise from the peduncle at maturity. The flesh is mostly green or white, occasionally orange, and not aromatic.

The Flexuosus group is made up of the snake or serpent melon and the Armenian cucumber. The fruit are quite long, thin, ribbed, and often curled irregularly.

The Conomon group comprises the oriental pickling melon. This fruit is smooth, cylindrical, and may be green, white, or striped. The flesh is white and can taste either sweet or bland.

The Dudaim group includes mango melon, pomegranate melon, and Queen Anne’s melon. The fruit are small, round to oval, and light green, yellow, or striped. The flesh is firm and yellowish-white in color.

The Mormordica group is made up of the phoot and snap melon. The fruit are oval or cylindrical with smooth skin that cracks as the fruit matures.

Plant and Fruit Morphology

Cucumber, melon, and watermelon plants share many characteristics but also differ in important ways. As a group they are frost-sensitive annuals with trailing, tendril-bearing vines. The plants are mostly monoecious, the flowers are insect-pollinated, and the fruits are variously shaped, many-seeded berries.

Cucumber

Cucumber plants are annual and may be monoecious, andromonoecious, or gynoeceous. They have indeterminate trailing vines with angled, hairy stems bearing triangular-ovate, acute three-lobed leaves. Determinate types with compact plants have been developed for gardens. In monoecious types, staminate flowers appear first and are several times more abundant than pistillate flowers. Flowers occur at the nodes, staminate in clusters or singly close to the plant crown with only one flower of the cluster opening on a single day; pistillate flowers are borne singly on the main stem and lateral branches in monoecious types (Figure II.C.6.5) and singly or in clusters on the main stem and lateral branches on gynoeceous types. Pistillate flowers are identified easily by the large inferior ovary that is a miniature cucumber fruit. Both staminate and pistillate flowers are large (2 to 3 centimeters [cm] in diameter) with a yellow, showy five-parted corolla. Fruits of commercial types are



Figure II.C.6.2. Casaba melon.

Figure II.C.6.3. Juan Canary melon.



cylindrical and green when consumed at the immature, edible stage (Figure II.C.6.6). The fruit surface is interrupted with tubercle-bearing white or black spines. White spines are typical of fruit used for fresh consumption that, if allowed to attain maturity, will be yellow, whereas black-spined fruit is often used for processing (pickles) and is orange at maturity.

Seedless or parthenocarpic cucumbers are another distinctive type. The plants are gynoeceous with a fruit borne at each axil (Figure II.C.6.7). They are grown on a trellis in protected, screened culture to prevent

bees from introducing foreign pollen, which would cause seeds to develop. Fruits are long, straight, smooth, thin-skinned, and medium to dark-green in color. A slightly restricted "neck" at the stem end of the fruit serves to readily identify this unique type. Cucumber fruit destined for fresh markets has a length/diameter ratio of about 4:1; that used for pickle production has a ratio of about 2:5, whereas parthenocarpic fruit have a ratio of about 6:1. Seeds are about 8 millimeters (mm) long, oval, and white (Lower and Edwards 1986: 173-81).



Figure II.C.6.4. Santa Claus melon.



Figure II.C.6.5. Pistillate and staminate cucumber flowers (Courtesy of National Garden Bureau, Downers Grove, Ill.)

West India Gherkin

These plants are annual, monoecious climbing vines with flowers, leaves, tendrils, and fruit smaller than those of cucumber. Fruits, which are spiny, yellow, oval, and about 5 cm long, are eaten fresh, cooked, or pickled. The plant may self-seed, escape from cultivation, and become an aggressive weed.

Melon

Melons are mostly andromonoecious and have annual trailing vines with nearly round stems bearing tendrils and circular to oval leaves with shallow lobes. Stami-

nate flowers are borne in axillary clusters on the main stem, and perfect flowers are borne at the first node of lateral branches. Fruits vary in size, shape, rind characteristics, and flesh color depending on variety. Fruit quality is related to external appearance, thick, well-colored interior flesh with high (>10 percent) soluble solids, and a pleasant aroma and taste (Maynard and Elmstrom 1991: 229). It is a common misconception that poor-quality melon fruit results from cross-pollination with cucumber because these species are incompatible. Rather, the poor-quality melon fruit sometimes encountered is due to unfavorable weather or grow-

Figure II.C.6.6. Cucumbers: commercial fresh-market type (right), pickling type (left), lemon cucumber (foreground), Armenian cucumber, which is a melon, *C. melo* (background). (Courtesy of National Garden Bureau, Downers Grove, Ill.)



ing conditions that restrict photosynthetic activity and, thereby, sugar content of the fruit. Seeds are cream-colored, oval, and on average 10 mm long.

Watermelon

These plants are monoecious, annual, and have trailing thin and angular vines that bear pinnatifid leaves. Flowers are solitary in leaf axils. Staminate flowers appear first and greatly outnumber pistillate flowers. The flowers are pollinated mostly by honeybees. Fruit may range in size from about 1 kilogram (kg) to as much as 100 kg, but ordinary cultivated types are 3 to 13 kg. Shape varies from round to oval to elongated. Coloration of the rind may be light green, often termed gray, to very dark green, appearing to be almost black (Figure II.C.6.8). In addition, the rind may have stripes of various designs that are typical of a variety or type; thus the terms "Jubilee-type stripe" or "Allsweet-type stripe" are used to identify various patterns. Seed color and size is variable. The tendency in varietal development is to strive for seeds that are small (but vigorous enough for germination under unfavorable conditions) and that are dark-colored rather than white – the latter are associated with immaturity. Flesh may be white, green, yellow, orange, pink, or red. Consumers in developed countries demand red- or deep pink-fleshed watermelons, although yellow-fleshed ones are grown in home gardens and, to a limited extent, commercially (Mohr 1986).

Seedless watermelon. Each fruit of standard-seeded watermelon varieties may contain as many as 1,000 seeds (Figure II.C.6.9) and their presence throughout the flesh makes removal difficult.



Figure II.C.6.7. Gynoecious, parthenocarpic greenhouse cucumbers.

Figure II.C.6.8. Variation in watermelon fruit size, shape, and color and flesh color. (Courtesy of National Garden Bureau, Downers Grove, Ill.)





Figure II.C.6.9. Seedless watermelon (center) with seeded watermelon (left and right).

Hybrid seedless (triploid) watermelons have been grown for over 40 years in the United States. However, only recently have improved varieties, aggressive marketing, and increased consumer demand created a rapidly expanding market for them. The seedless condition is actually sterility resulting from a cross between two plants of incompatible chromosome complements. The normal chromosome number in most living organisms is referred to as $2n$. Seedless watermelons are produced on highly sterile triploid ($3n$) plants, which result from crossing a normal diploid ($2n$) plant with a tetraploid ($4n$). The tetraploid is used as the female or seed parent and the diploid is the male or pollen parent. Since the tetraploid seed parent produces only 5 to 10 percent as many seeds as a normal diploid plant, seed cost is 10 to 100 times more than that of standard, open-pollinated varieties and 5 to 10 times that of hybrid diploid watermelon varieties.

Tetraploid lines, usually developed by treating diploid plants with a chemical called colchicine, normally have a light, medium, or dark-green rind without stripes. By contrast, the diploid pollen parent almost always has a fruit with a striped rind. The resulting hybrid triploid seedless melon will inherit the striped pattern, though growers may occasionally find a nonstriped fruit in fields of striped seedless watermelons, the result of accidental self-pollination of the tetraploid seed parent during triploid seed production. The amount of tetraploid contamination depends upon the methods and care employed in triploid seed production. Sterile triploid plants normally do not produce viable seed. However, small, white rudimentary seeds or seed coats, which are eaten along with the fruits as in cucumber, develop within the fruit. The number and size of these rudi-

mentary seeds vary with the variety. An occasional dark, hard, viable seed is found in triploid melons. Seedless watermelons can be grown successfully in areas where conventional seeded varieties are produced, although they require some very unique cultural practices for successful production (Maynard 1996: 1-2). With proper care, such watermelons have a longer shelf life than seeded counterparts. This may be due to the fact that flesh breakdown occurs in the vicinity of seeds, which are absent in seedless melons.

Citron

The citron plants resemble those of watermelon except that their leaves are broader and less pinnate. The fruits also resemble watermelon externally, but the rind is quite hard and the flesh is white to light green and may be quite bitter. Because fruit rinds are used to make pickles and are also candied, the citron is also called a "preserving melon." Plants escaped from cultivation may prove to be aggressive weeds in crop fields.

History and Ethnography of Production and Consumption

Relatively little research literature in cultural anthropology, archaeology, or social history focuses specifically on the species of cultivated cucurbits under consideration here. Indeed, some classic as well as recent important texts on the origins of agriculture make no mention of them (Reed 1977; Smith 1995). There are at least four reasons for this lacuna. First, these cultigens are not part of the complex carbohydrate "cores" of the diets found in centers of state formation (see Mintz 1996) and thus have not received the same attention as other staple food crops. Second, the primary centers of

domestication for both melon and watermelon are in sub-Saharan Africa, where the exact timing, locations, and processes of domestication are still poorly understood (see Cowan and Watson 1992). Third, some researchers suggest that "cucurbits are usually poorly preserved among archaeological remains. The features considered most indicative of domestication are characteristics of the peduncle (stem), which is rarely preserved. The earliest remains are seed specimens, which often occur in extremely low frequencies because they are likely to have been consumed" (McClung de Tapia 1992: 153). Finally, the ethnographic record contains limited data on the production and consumption of these crops (with a few notable exceptions), reflecting their secondary significance both materially and symbolically in most human societies.

Cucumber

Cucumbers are generally believed to have originated in India, and archaeological and linguistic evidence suggests that they have been cultivated throughout western Asia for at least 3,000 years (Hedrick 1919: 208; Whitaker and Davis 1962: 2-3; Sauer 1993: 45; Robinson and Decker-Walters 1997: 62). From India, the cucumber spread to Greece and Italy - where the crop was significant in the Roman Empire - and slightly later to China and southern Russia. In classical Rome, Pliny reported greenhouse production of cucumbers by the first century, and the Emperor Tiberius was said to have had them at his table throughout the year (Sauer 1993: 46). Cucumbers probably were diffused into the rest of Europe by the Romans and later throughout the New World via colonialism and indigenous trade networks. The earliest records of their cultivation appear in France by the ninth century, Great Britain by the fourteenth century, the Caribbean at the end of the fifteenth century, and North America by the middle of the sixteenth century (Hedrick 1919: 208).

Colonial encounters between Europeans and Native Americans resulted in the rapid diffusion of cucumbers throughout North America. The Spanish began growing them in Hispaniola by 1494, and less than a century later European explorers were noting that a wide range of Native American peoples from Montreal to New York, Virginia, and Florida were cultivating them, along with a large variety of other crops including maize, beans, squash, pumpkins, and gourds. By the seventeenth century, Native American groups on the Great Plains were also cultivating cucumbers - this in a region where the Spanish had been particularly significant in the diffusion of horses and guns, as well as Old World cultigens such as watermelons and cucumbers (see Wolf 1982).

Like other cucurbits, cucumbers have a wide range of consumption uses cross-culturally. They are generally eaten fresh or pickled and are particularly important in the diets of people living in Russia and East, South, and Southeast Asia, where they may also be served as a fresh or cooked vegetable. In India, the

fruits are used in the preparation of chutney and curries. Cucumber seeds, young leaves, and cooked stems are also consumed in some parts of Asia.

In addition, since at least the nineteenth century, cucumbers have been used in the production of a large variety of cosmetics, including fragrances, body lotions, shampoos, and soaps (Robinson and Decker-Walters 1997: 63; Rubatzky and Yamaguchi 1997: 585).

Melon

Melon is generally thought to have originated in western Africa (Zeven and Zhukovsky 1975: 30; Bailey 1976: 342; Purseglove 1976: 294; Whitaker and Bemis 1976: 67), with China or India as possible secondary centers of diversity. Wild melons growing in natural habitats have been reported in desert and savanna zones of Africa, Arabia, southwestern Asia, and Australia. As Jonathan Sauer notes, it is unclear where melon was domesticated and "it is conceivable that it was independently domesticated from different wild populations in Africa and southwestern Asia" (Sauer 1993: 44). Melon was an important food crop in ancient China, where archaeological data suggest that it has been cultivated for over 5,000 years (Robinson and Decker-Walters 1997: 23). Archaeological evidence also suggests that melon was cultivated in Iran some 5,000 years ago and in Greece and Egypt about 4,000 years ago (Zohary and Hopf 1988). Given the fruit's probable African origin, this evidence points to a very early date for the first domestication of melon. Tropical forest swidden systems in Africa typically have yams or manioc as dominant staple food crops with melons among the numerous and multiple secondary crops (Harris 1976: 318).

As with cucumbers, melons were cultivated in the Roman Empire and diffused throughout Europe by the Middle Ages where the "variety and quality of melon cultivars were evidently greatly increased by selection in Medieval gardens" (Sauer 1993: 44). As with cucumbers and watermelons, melons were introduced to the New World by Spanish colonial settlers in the late fifteenth and early sixteenth centuries and subsequently spread very rapidly among Native American horticultural groups. Later during the eighteenth century they reached the Pacific Islanders via British explorers.

Ralf Norrman and Jon Haarberg (1980) explore the semiotic role of cucurbits in Western literature and culture and extend this analysis to selected non-Western cultural contexts. Focusing on melons, watermelons, and cucumbers (as well as other domesticated cucurbits), these authors note that cucurbits generally have deep, profound, and complex multivocal symbolic associations with sex and sexuality, fertility, vitality, moisture, abundance, opulence, luxury, gluttony, creative power, rapid growth, and sudden death. More specifically, they note that melons are highly associated with status in colder climate European societies because historically they were "seasonal, expensive and scarce, with all the symbolic development that a

commodity with such characteristics usually goes through" (Norrman and Haarberg 1980: 16). Cucurbits also appear frequently in non-Western cosmologies, for example, "in Burmese and Laotian mythology, the creation of man started from a cucurbit" (Norrman and Haarberg 1980: 26). As with other key symbols marked by binary oppositions, symbolic meanings attached to cucurbits can also be employed to convey a broad variety of negative symbolic associations along race, class, and gender lines.

Melon has a large number of different cultivars and a range of cross-cultural consumption uses parallel to the other species of cucurbits discussed in this chapter. Fruits are typically eaten uncooked, although they may also be cooked or pickled in some Asian cuisines. The seeds of some cultivars are roasted and consumed in parts of India. Dried and ground melon seeds are used as food in some African societies. Melon fruits, roots, leaves, and seeds play important roles in the treatment of a wide range of health problems in Chinese traditional medicine (Robinson and Decker-Walters 1997: 69–70).

Watermelon

Watermelons, which were originally domesticated in central and southern Africa (Whitaker and Davis 1962: 2; Robinson and Decker-Walters 1997: 85), are an important part of the "most widespread and characteristic African agricultural complex adapted to savanna zones" in that they are not only a food plant but also a vital source of water in arid regions (Harlan, de Wet, and Stemler 1976; Harlan 1992: 64). Indeed, V. R. Rubatzky and M. Yamaguchi (1997: 603) refer to watermelons as "botanical canteens." In a number of traditional African cuisines, the seeds (rich in edible oils and protein) and flesh are used in cooking. Watermelon emerged as an important cultivar in northern Africa and southwestern Asia prior to 6,000 years ago (Robinson and Decker-Walters 1997: 24). Archaeological data suggest that they were cultivated in ancient Egypt more than 5,000 years ago, where representations of watermelons appeared on wall paintings and watermelon seeds and leaves were deposited in Egyptian tombs (Ficklen 1984: 8).

From their African origins, watermelons spread via trade routes throughout much of the world, reaching India by 800 and China by 1100. In both of these countries, as in Africa, the seeds are eaten and crushed for their edible oils. Watermelons became widely distributed along Mediterranean trade routes and were introduced into southern Europe by the Moorish conquerors of Spain, who left evidence of watermelon cultivation at Cordoba in 961 and Seville in 1158 (Watson 1983). Sauer notes that "watermelons spread slowly into other parts of Europe, perhaps largely because the summers are not generally hot enough for good yields. However,

they began appearing in European herbals before 1600, and by 1625, the species was widely planted in Europe as a minor garden crop" (Sauer 1993: 42). Their first recorded appearance in Great Britain dates to 1597.

Watermelons reached the New World with European colonists and African slaves. Spanish settlers were producing watermelons in Florida by 1576, and by 1650 they were common in Panama, Peru, and Brazil, as well as in British and Dutch colonies throughout the New World (Sauer 1993: 43). The first recorded cultivation in British colonial North America dates to 1629 in Massachusetts (Hedrick 1919: 172).

Like cucumbers and melons, watermelons spread very rapidly among Native American groups. Prior to the beginning of the seventeenth century, they were being grown by tribes in the Ocmulgee region of Georgia, the Conchos nation of the Rio Grande valley, the Zuni and other Pueblo peoples of the Southwest, as well as by the Huron of eastern Canada and groups from the Great Lakes region (Blake 1981). By the mid-seventeenth century, Native Americans were cultivating them in Florida and the Mississippi valley, and in the eighteenth and early nineteenth centuries the western Apache of east-central and southeastern Arizona were producing maize and European-introduced crops including watermelons as they combined small-scale horticulture with hunting and gathering in a low rainfall environment (Minnis 1992: 130–1). This fact is ethnographically significant because other transitional foraging-farming groups, such as the San people of the Kalahari Desert of southern Africa, have parallel subsistence practices involving watermelons. Watermelons and melons were also rapidly adopted by Pacific Islanders in Hawaii and elsewhere as soon as the seeds were introduced by Captain James Cook (1778) and other European explorers (Neal 1965).

In the cultural history of the United States, Thomas Jefferson was an enthusiastic grower of watermelons at his Monticello estate, Henry David Thoreau proudly grew large and juicy watermelons in Concord, Massachusetts, and Mark Twain wrote in *Pudd'n'head Wilson*: "The true southern watermelon is a boon apart and not to be mentioned with commoner things. It is chief of this world's luxuries, king by the grace of God over all the fruits of the earth. When one has tasted it, he knows what the angels eat." Ellen Ficklen has documented the important role of watermelons in American popular culture in numerous areas including folk art, literature, advertising and merchandising, and the large number of annual summer watermelon festivals throughout the country with "parades, watermelon-eating contests, seed spitting contests, watermelon queens, sports events, and plenty of food and music" (1984: 25).

Growing and exhibiting large watermelons is an

active pastime in some rural areas of the southern United States. Closely guarded family “secrets” for producing large watermelons and seeds from previous large fruit are carefully maintained. According to *The Guinness Book of Records*, the largest recorded watermelon in the United States was grown by B. Carson of Arrington, Tennessee, in 1990 and weighed a phenomenal 119 kg (Young 1997: 413).

African slaves also widely dispersed watermelon seeds in eastern North America, the circum-Caribbean, and Brazil. In the southern United States – where soil and climate conditions were optimal for watermelon cultivation – this crop ultimately became stereotypically, and often negatively, associated with rural African-Americans (see Norrman and Haarberg 1980: 67–70). Watermelons have subsequently figured as key symbols in the iconography of racism in the United States as seen during African-American protest marches in Bensonhurst, Brooklyn, in 1989, where marchers were greeted by Italian-American community residents shouting racial slurs and holding up watermelons.

In the ethnographic record of cultural anthropology, watermelons have perhaps figured most extensively in discussions of foragers and agro-pastoralists of the Kalahari Desert in southern Africa. As early as the 1850s, explorer David Livingstone described vast tracts of watermelons growing in the region. The anthropologist Richard Lee notes that watermelons, in domestic, wild, and feral varieties, constitute one of the most widespread and abundant plant species growing in the central Kalahari Desert. They are easily collected by foraging peoples and “the whole melon is brought back to camp and may be cut into slices for distribution. The melon itself may be halved and used as a cup, while the pulp is pulverized with the blunt end of a digging stick. The seeds may be roasted and eaten as well” (Lee 1979: 488).

Watermelons are among the most popular cultivars for forager-farmers in the Kalahari for the following reasons: “First, they provide a source of water; second, they are relatively drought-resistant, especially when compared to seed crops like sorghum and maize; and third, dried melons are an article of food for both humans and livestock and, after they have been cut into strips and hung on thorn trees to dry, they are easy to store” (Hitchcock and Ebert 1984: 343).

Elizabeth Cashdan emphasizes the point that “normally, when one thinks of agriculture one thinks of food resources, but . . . where the dominant factor governing mobility is the availability of moisture, it is appropriate that agriculture should be used to produce a storable form of moisture” (Cashdan 1984: 316). This cultivated water supply allows some Kalahari Desert groups to remain sedentary during both rainy and dry seasons, and watermelons are often stored in large quantities by these societies (Cashdan 1984: 321).

The collection of watermelons by foragers and their incipient domestication by such groups yields insights into probable scenarios for domestication. R. W. Robinson and D. S. Decker-Walters suggest a general process for cucurbits that has a plausible fit with the history and ethnography of watermelons in the Kalahari Desert:

Aboriginal plant gatherers were probably attracted to some of these products, particularly the relatively large, long-keeping and sometime showy fruits. After fruits were taken back to camp, seeds that were purposely discarded, accidentally dropped or partially digested found new life on rubbish heaps, settlement edges or other disturbed areas within the camp. Eventually, recognition of the value of the resident cucurbits led to their tolerance, horticultural care and further exploitation. Finally seeds . . . were carried by and exchanged among migrating bands of these incipient cultivators, gradually turning the earliest cultivated cucurbits into domesticated crops. (Robinson and Decker-Walters 1997: 23)

Such a process of domestication is somewhat different from those analyzed for cereal grains, where early transitional forager-farmers exploited densely concentrated stands of the wild ancestors of later domesticated varieties.

Cross-cultural uses of watermelon are quite varied. They are primarily consumed fresh for their sweet and juicy fruits and are often eaten as desserts. In some African cuisines, however, they are served as a cooked vegetable. The rind may be consumed in pickled or candied form. In parts of the former Soviet Union and elsewhere watermelon juice is fermented into an alcoholic beverage. Roasted seeds of this crop are eaten throughout Asia and the Middle East, and watermelon seeds are ground into flour and baked as bread in some parts of India. In addition, watermelons are also sometimes used as feed for livestock (Robinson and Decker-Walters 1997: 24–7, 85; Rubatzky and Yamaguchi 1997: 603).

Variety Improvement

Cucumber

Early cucumber varieties used in the United States were selections of those originally brought from Europe. American-originated varieties such as ‘Arlington White Spine’, ‘Boston Pickling’, and ‘Chicago Pickling’ were developed in the late nineteenth century. Cucumber is prone to a large number of potentially devastating diseases, and its rapid trailing growth makes chemical control of foliar and fruit diseases quite difficult. As a result, interest in the development of genetic disease tolerance has long been the focus of plant breeding efforts and has met with great success:

Tolerance to at least nine diseases has been incorporated into a single genotype. The first monoecious hybrid, 'Burpee Hybrid', was made available in 1945. Although seed costs were higher, multiple advantages of hybrids were soon recognized. Commercial companies built large research staffs to develop hybrids that provided proprietary exclusivity in those species where appropriate. Gynoeious hybrids made their appearance in 1962 when 'Spartan Dawn' was introduced. This all-female characteristic has since been exploited in both pickling and fresh-market types (Wehner and Robinson 1991: 1-3).

Melon

An 1806 catalog lists 13 distinct melon sorts derived from European sources (Tapley, Enzie, and Van Eseltine 1937: 60). Management of plant diseases in melon presents the same difficulties as with cucumbers. Accordingly, incorporation of disease tolerance into commercial types has been a major objective of plant breeders. One type, 'PMR 45', developed by the U.S. Department of Agriculture and the University of California in 1937, represented an enormous contribution because it provided resistance to powdery mildew (*Erysiphe cichoracearum*), which was the most devastating disease of melons in the arid western United States. This variety and its descendants dominated the U.S. market for about 40 years (Whitaker and Davis 1962: 57-9). Hybrids, which now predominate in the *Cantalupensis* group, began to appear in the mid-1950s with the introduction of 'Burpee Hybrid', 'Harper Hybrid', and others (Minges 1972: 69, 71).

Watermelon

Tolerance to fusarium wilt (*Fusarium oxysporum* f. sp. *niveum*) and anthracnose (*Colletotrichum orbiculare*), which was a prime objective of watermelon breeding programs, was achieved with the development of three varieties that dominated commercial production for almost four decades. 'Charleston Gray' was developed by C. F. Andrus of the U.S. Department of Agriculture in 1954, 'Crimson Sweet' by C. V. Hall of Kansas State University in 1964, and 'Jubilee' by J. M. Crall at the University of Florida in 1963 (Figure II.C.6.10). These varieties are no longer used to any extent, having been replaced by hybrids of the Allsweet and blocky Crimson Sweet types because of superior quality, high yields, and an attractive rind pattern. In Japan and other parts of Asia, watermelon varieties in use are susceptible to fusarium wilt, so they are grafted (Figure II.C.6.11) onto resistant root stocks (Lee 1994). In addition to diploid hybrids, triploid (seedless) hybrids are expected to dominate the watermelon market in the near future.

Production, Consumption, and Nutritional Composition

Production

Cucumber. As Table II.C.6.1 indicates, well over half of world cucumber and gherkin production occurs in Asia (the term "gherkin" is used here to denote small cucumber, rather than bur or West India gherkin). Though significant production also occurs in Europe and in North and Central



Figure II.C.6.10. 'Jubilee' watermelon developed by J. M. Crall, University of Florida, in 1963.

Figure II.C.6.11. In Japan, watermelon seedlings are grafted by machine onto Fusarium-resistant rootstocks.



America, China accounts for about 40 percent of world production. Other Asian countries with high cucumber production are Iran, Turkey, Japan, Uzbekistan, and Iraq. Only the United States, Ukraine, the Netherlands, and Poland are world leaders outside of Asia in cucumber production. Yields in the leading producing countries range from 8.6 tons per hectare (ha) in Iraq to 500 tons/ha in the Netherlands. The extraordinary yields in the Netherlands are because of protected culture of parthenocarpic types (United Nations 1996: 134-5).

Table II.C.6.1. *World cucumber and gherkin production, 1995*

Location	Area (ha × 10 ³)	Yield (t × ha ⁻¹)	Production (t × 10 ³)
World	1,200	16.1	19,353
Africa	23	17.0	388
North and Central America	106	13.8	1,462
South America	4	16.8	67
Asia	780	17.1	13,372
Europe	90	27.2	2,434
Oceania	1	15.8	21
Leading countries			
China	468 ^a	17.2	8,042 ^a
Iran	72 ^a	17.4	1,250 ^a
Turkey	41 ^a	28.0	1,150 ^a
United States	71 ^a	14.1	992 ^a
Japan	19 ^a	45.6	866 ^a
Ukraine	61	11.0	669
Netherlands	1 ^a	500.0	500 ^a
Poland	34 ^a	10.9	370 ^a
Uzbekistan	35 ^a	10.0	350 ^a
Iraq	40 ^a	8.6	346 ^a

^aEstimated.

Source: United Nations (1996), pp. 134-5.

Melon. As with cucumber, Asia produces more than half of the world's melon crop (Table II.C.6.2). Whereas Europe, North and Central America, and Africa are important world production centers, China produces about 25 percent of the world's crop. Turkey and Iran are also leading melon-producing countries. Yields in the leading countries range from 13.0 tons/ha in Mexico to 26.9 tons/ha in China (United Nations 1996: 122-3). In Japan, melons are usually grown in greenhouses. The very best ones are sold to be used as special gifts. Prices shown (Figure II.C.6.12) are roughly U.S. \$50, \$60, and \$70 each.

Table II.C.6.2. *World cantaloupe and other melon production, 1995*

Location	Area (ha × 10 ³)	Yield (t × ha ⁻¹)	Production (t × 10 ³)
World	823	17.0	14,018
Africa	61	16.9	1,024
North and Central America	116	15.6	1,805
South America	41	7.6	309
Asia	460	18.3	8,422
Europe	142	16.8	2,382
Oceania	4	21.1	76
Leading countries			
China	130 ^a	26.9	3,492 ^a
Turkey	110 ^a	16.4	1,800 ^a
Iran	88 ^a	13.8	1,215 ^a
United States	42 ^a	20.5	859 ^a
Spain	43 ^a	19.0	820 ^a
Romania	50 ^a	13.6	680 ^a
Mexico	50 ^a	13.0	650 ^a
Egypt	25 ^a	18.4	480 ^a
Morocco	25 ^a	16.7	415 ^a
Japan	18 ^a	22.3	390 ^a

^aEstimated.

Source: United Nations (1996), pp. 122-3.



Figure II.C.6.12. Melons for sale as special gifts in Kyoto, Japan.

Table II.C.6.3. World watermelon production, 1995

Location	Area (ha $\times 10^3$)	Yield (t $\times$ ha $^{-1}$)	Production (t $\times 10^3$)
World	1,823	16.3	29,656
Africa	120	16.3	1,956
North and Central America	134	17.9	2,394
South America	131	8.8	1,153
Asia	905	19.3	17,502
Europe	122	21.8	2,652
Oceania	5	17.0	80
Leading countries			
China	359 ^a	18.6	6,696 ^a
Turkey	135 ^a	28.8	3,600 ^a
Iran	145 ^a	18.3	2,650 ^a
United States	86 ^a	21.1	1,808 ^a
Korea Republic	38 ^a	23.7	900 ^a
Georgia	60 ^a	13.3	800 ^a
Egypt	34 ^a	21.2	720 ^a
Uzbekistan	62 ^a	11.3	700 ^a
Japan	22 ^a	30.4	655 ^a
Moldova Republic	50 ^a	13.0	650 ^a

^aEstimated.

Source: United Nations (1996), pp. 146–7.

Watermelon. Asia produces about 60 percent of the world's watermelons with major production in China (23 percent), Turkey (12 percent), Iran (9 percent), Korea Republic (3 percent), Georgia (3 percent), Uzbekistan (2 percent), and Japan (2 percent) (Table II.C.6.3). Yields in the major producing countries range from 11.3 tons/ha in Uzbekistan to 30.4 tons/ha in Japan (Figure II.C.6.13), where much of the production is in protected culture (United Nations 1996: 146–7).



Figure II.C.6.13. Low, supported row covers for watermelon production in Daiei, Japan.

Consumption and Nutritional Composition

Cucurbits, as previously discussed in this chapter, are an important part of the diet in the United States (Table II.C.6.4), where the annual consumption of watermelon, melon, and cucumber amounts to just over 17 kg per person (USDA 1996). Cucurbit fruits are high in moisture and low in fat, which makes them popular with consumers interested in healthy diets (Table II.C.6.5). Those with orange flesh like muskmelon and winter squash are excellent sources of vitamin A. Orange-fleshed cucumbers have been developed recently from crosses between United States pickling cucumber varieties and the orange-fruited "Xishuangbanna" cucumber from the People's Republic of China. The provitamin A carotene content of these cucumbers is equivalent to other orange-fleshed cucurbits (Simon and Navazio 1997). Moderate amounts of essential

inorganic elements and other vitamins are provided by the cucurbit fruit. Aside from the low fat content and high vitamin A content of some cucurbit fruits, their principal value in the diet of people living in developed countries is in their unique colors, shapes, flavors, and adaptability to various cuisines.

The internal quality of watermelon fruit is a function of flesh color and texture, freedom from defects, sweetness, and optimum maturity. Unfortunately, these criteria cannot, as a rule, be assessed without cutting the melon. So many watermelons of inferior or marginal quality have been marketed that consumers have increasingly lost confidence in the product. The current supermarket practice of preparing cut and sectioned watermelon provides at least partial assurance of quality to the purchaser, but no indication of sweetness. In Japan, the quality of whole watermelon fruit is assessed by nuclear magnetic resonance (NMR) before marketing. Soluble solids and flesh integrity can be determined nondestructively in seconds (Figure II.C.6.14). As mentioned, because of their exceptional quality, such watermelons can be sold locally for the equivalent of about U.S. \$50-\$70 (Figure II.C.6.15).

In contrast to the composition of the pulp, watermelon seeds, which are used for food in various parts of the world, are low in moisture and high in carbohydrates, fats, and protein. Varieties with very large seeds have been developed especially for use as food in China, where more than 200,000 tons are produced annually on 140,000 ha land (Zhang 1996).

David Maynard
Donald N. Maynard

Table II.C.6.4. *Per capita consumption of cucumbers, melons, and watermelons in the United States, 1996*

Vegetable	Consumption (kg)
Cucumber - fresh	2.54
Cucumber - processed	2.18
Honeydew melon	1.13
Muskmelon	4.26
Watermelon	7.26
All vegetables - fresh	90.81
All vegetables - processed	106.86

Source: USDA (1996), VGS-269.

Table II.C.6.5 *Nutritional composition of some cucurbits; amounts per 100 g edible portion*

Nutrient	Cucumber (slicing)	Cucumber (pickling)	West India gherkin	Casaba melon	Honeydew melon	Musk-melon	Watermelon (fruit)	Watermelon (seed)	Summer squash	Winter squash
Water (%)	96	96	93	92	90	90	93	5.7	94	89
Energy (kcal)	13	12	17	26	35	35	26	567	20	37
Protein (g)	0.5	0.7	1.4	0.9	0.5	0.9	0.5	25.8	1.2	1.5
Fat (g)	0.1	0.1	0.3	0.1	0.1	0.3	0.2	49.7	0.2	0.2
Carbohydrate (g)	2.9	2.4	2.0	6.2	9.2	8.4	6.4	15.1	4.4	8.8
Fiber (g)	0.6	0.6	0.6	0.5	0.6	0.4	-	4.0	0.6	1.4
Ca (mg)	14	13	26	5	6	11	7	53	20	31
P (mg)	17	24	38	7	10	17	10	-	35	32
Fe (mg)	0.3	0.6	0.6	0.4	0.1	0.2	0.5	-	0.5	0.6
Na (mg)	2	6	6	12	10	9	1	-	2	4
K (mg)	149	190	290	210	271	309	100	-	195	350
Vitamin A (IU)	45	270	270	30	40	3,224	590	-	196	4,060
Thiamine (mg)	0.03	0.04	0.1	0.06	0.08	0.04	0.03	0.1	0.06	0.10
Riboflavin (mg)	0.02	0.2	0.04	0.02	0.02	0.02	0.03	0.12	0.04	0.03
Niacin (mg)	0.30	0.4	0.4	0.40	0.60	0.57	0.20	1.4	0.55	0.80
Ascorbic acid (mg)	4.7	19.0	51.0	16.0	24.8	42.2	7.0	-	14.8	12.3
Vitamin B ₆	0.05	0.4	0.4	-	0.06	0.12	-	1.4	0.11	0.08

Sources: Gebhardt, Cutrufelli, and Matthews (1982), Haytowitz and Matthews (1984), and Rubatzky and Yamaguchi (1997).



Figure II.C.6.14. NMR watermelon quality determination in Japan.



Figure II.C.6.15. Watermelon for sale in Japan at U.S. \$50.

Bibliography

- Bailey, Liberty Hyde. 1976. *Hortus third*. New York.
- Bates, David M., Richard W. Robinson, and Charles Jeffrey, eds. 1990. *Biology and utilization of the cucurbitaceae*. Ithaca, N.Y.
- Blake, L. W. 1981. Early acceptance of watermelons by Indians in the United States. *Journal of Ethnobiology* 1: 193-9.
- Cashdan, Elizabeth. 1984. The effects of food production on mobility in the Central Kalahari. In *From hunters to farmers: The causes and consequences of food production*, ed. J. Desmond Clark and Steven A. Brandt, 311-27. Berkeley, Calif.
- Cowan, C. Wesley, and Patty Jo Watson. 1992. Some concluding remarks. In *The origins of agriculture: An international perspective*, ed. C. Wesley Cowan and Patty Jo Watson, 207-12. Washington, D.C.
- Ficklen, Ellen. 1984. *Watermelon*. Washington, D.C.
- Gebhardt, S. E., R. Cutrufelli, and R. H. Matthews. 1982. *Composition of foods, fruits and fruit juices - raw, processed, prepared*. U.S. Department of Agriculture Handbook, 8-9.
- Gómez-Guillamón, M. L., ed. 1996. *Cucurbits towards 2000*. Malaga, Spain.
- Harlan, Jack R. 1992. Indigenous African agriculture. In *The origins of agriculture: An international perspective*, ed. C. Wesley Cowan and Patty Jo Watson, 59-70. Washington, D.C.
- Harlan, Jack R., J. M. J. de Wet, and Ann Stemler. 1976. Plant domestication and indigenous African agriculture. In *Origins of African plant domestication*, ed. Jack Harlan, Jan M. J. de Wet, and Ann B. L. Stemler, 3-19. The Hague.
- Harris, David R. 1976. Traditional systems of plant food production and the origins of agriculture in West Africa. In *Origins of African plant domestication*, ed. Jack Harlan, Jan M. J. de Wet, and Ann B. L. Stemler, 311-56. The Hague.

- Haytowitz, D. B., and R. H. Matthews. 1984. *Composition of foods, vegetables and vegetable products - raw, processed, prepared*. U.S. Department of Agriculture Handbook, 8-11.
- Hedrick, U. P. 1919. Sturtevant's notes on cultivated plants. *New York Department of Agriculture Annual Report* 27 (2, II): 1-686. New York.
- Hitchcock, Robert K., and James I. Ebert. 1984. Foraging and food production among Kalahari hunter/gatherers. In *From hunters to farmers: The causes and consequences of food production*, ed. J. Desmond Clark and Steven A. Brandt, 328-48. Berkeley, Calif.
- Jeffrey, Charles. 1990a. An outline classification of the Cucurbitaceae. In *Biology and utilization of the Cucurbitaceae*, ed. D. M. Bates, R. W. Robinson, and C. Jeffrey, 449-63. Ithaca, N.Y.
- 1990b. Systematics of the Cucurbitaceae: An overview. In *Biology and utilization of the Cucurbitaceae*, ed. D. M. Bates, R. W. Robinson, and C. Jeffrey, 3-9. Ithaca, N.Y.
- Lee, Jung-Myung. 1994. Cultivation of grafted vegetables I. Current status, grafting methods, and benefits. *HortScience* 29: 235-9.
- Lee, Richard B. 1979. *The Kung San: Men, women, and work in a foraging society*. New York.
- Lester, G. E., and J. R. Dunlap, ed. 1994. *Proceedings of cucurbitaceae 94*. Edinburg, Tex.
- Lower, R. L., and M. O. Edwards. 1986. Cucumber breeding. In *Breeding vegetable crops*, ed. M. J. Bassett, 173-207. Westport, Conn.
- Maynard, D. N. 1996. *Growing seedless watermelons*. University of Florida, Gainesville.
- Maynard, D. N., and G. W. Elmstrom. 1991. Potential for western-type muskmelon production in central and southwest Florida. *Proceedings of the Florida State Horticultural Society* 104: 229-32.
- McClung de Tapia, Emily. 1992. The origins of agriculture in Mesoamerica and Central America. In *The origins of agriculture: An international perspective*, ed. C. Wesley Cowan and Patty Jo Watson, 143-72. Washington, D.C.
- Minges, P. A., ed. 1972. *Descriptive list of vegetable varieties*. Washington D.C. and St. Joseph, Mich.
- Minnis, Paul E. 1992. Earliest plant cultivation in the desert borderlands of North America. In *The origins of agriculture: An international perspective*, ed. C. Wesley Cowan and Patty Jo Watson, 121-41. Washington, D.C.
- Mintz, Sidney W. 1996. *Tasting food, tasting freedom: Excursions into eating, culture and the past*. Boston, Mass.
- Mohr, H. C. 1986. Watermelon breeding. In *Breeding vegetable crops*, ed. M. J. Bassett, 37-42. Westport, Conn.
- Munger, H. M., and R. W. Robinson. 1991. Nomenclature of *Cucumis melo* L. *Cucurbit Genetics Cooperative Report* 14: 43-4.
- Neal, M. C. 1965. *In gardens of Hawaii*. Honolulu.
- Ng, T. J., ed. 1996. *Cucurbit Genetics Cooperative Report*. College Park, Md.
- Norrman, Ralf, and Jon Haarberg. 1980. *Nature and languages: A semiotic study of cucurbits in literature*. London.
- Purseglove, J. W. 1976. The origins and migration of crops in tropical Africa. In *Origins of African plant domestication*, ed. Jack Harlan, Jan M. J. de Wet, and Ann B. L. Stemler, 291-309. The Hague.
- Reed, Charles A., ed. 1977. *Origins of agriculture*. The Hague.
- Robinson, R. W., and D. S. Decker-Walters. 1997. *Cucurbits*. New York.
- Rubatzky, V. R., and M. Yamaguchi. 1997. *World vegetables: Principles, production, and nutritive value*. New York.
- Sauer, Jonathan D. 1993. *Historical geography of crop plants: A select roster*. Boca Raton, Fla.
- Simon, P. W., and J. P. Navazio. 1997. Early orange mass 400, early orange mass 402, and late orange mass 404: High-carotene cucumber germplasm. *HortScience* 32: 144-5.
- Smith, Bruce D. 1995. *The emergence of agriculture*. New York.
- Tapley, W. T., W. D. Enzie, and G. P. Van Eseltine. 1937. *The vegetables of New York*. Vol. 1, Part IV, *The cucurbits*. New York State Agricultural Experiment Station, Geneva.
- Thomas, C. E., ed. 1989. *Proceedings of Cucurbitaceae 89*. Charleston, S.C.
- United Nations. 1996. *FAO production yearbook*. Rome.
- USDA (U.S. Department of Agriculture). 1996. *Vegetables and specialties: Situation and outlook*. VGS-269. Washington, D.C.
- Watson, A. M. 1983. *Agricultural innovation in the early Islamic world: The diffusion of crops and foraging techniques, 700-1100*. New York.
- Wehner, T. C., and R. W. Robinson. 1991. A brief history of the development of cucumber cultivars in the United States. *Cucurbit Genetics Cooperative Report* 14: 1-3.
- Whitaker, Thomas W., and W. P. Bemis. 1976. Cucurbits. In *Evolution of crop plants*, ed. N. W. Simmonds, 64-9. London.
- Whitaker, T. W., and G. N. Davis. 1962. *Cucurbits. Botany, cultivation, and utilization*. New York.
- Wolf, Eric R. 1982. *Europe and the people without history*. Berkeley, Calif.
- Young, M. C., ed. 1997. *The Guinness book of records*. New York.
- Zeven, A. C., and P. M. Zhukovsky. 1975. *Dictionary of cultivated plants and their centers of diversity*. Wageningen, the Netherlands.
- Zhang, J. 1996. Breeding and production of watermelon for edible seed in China. *Cucurbit Genetics Cooperative Report* 19: 66-7.
- Zohary, Daniel, and Maria Hopf. 1988. *Domestication of plants in the Old World: The origin and spread of cultivated plants in West Asia, Europe, and the Nile Valley*. Oxford.

H.C.7 Fungi

Definitions

Fungi are uninucleate or multinucleate, eukaryotic organisms with nuclei scattered in a walled and often septate mycelium (the vegetative part of a fungus). Nutrition is heterotrophic (at least one or more organic molecules required), and fungi usually obtain their nutrients by way of diffusion or active transport.

They lack chlorophyll but may have other pigments such as carotenoids, flavonoids, and so forth.

The true fungi, Eumycota, are grouped into five divisions:

1. Mastigomycotina (aquatic or zoospore-producing fungi) – unicellular or mycelial (coenocytic, no intercellular walls); motile, uni- or biflagellate zoospores during life cycle.
2. Zygomycotina – coenocytic mycelium; sexual state (teleomorph) spores are zygospores which may be absent; asexual state (anamorph) is predominant stage consisting of uni- or multispored sporangia.
3. Ascomycotina – mycelium unicellular to multicellular; regularly septate; asexual state often present; sexual state spores are ascospores formed inside an ascus (sac); no motile state.
4. Basidiomycotina – mycelium unicellular to multicellular, regularly septate; conidial asexual state common; sexual state and motile cells absent.
5. Deuteromycotina – unicellular to multicellular mycelia; regularly septate; conidial asexual state common; no sexual state; no motile cells (O'Donnell and Peterson 1992).

D. L. Hawksworth, B. C. Sutton, and G. A. Ainsworth (1983) have estimated that there are about 250,000 species of fungi, of which Mastigomycotina composes 1.8 percent, Zygomycotina 1.2 percent, Ascomycotina about 45 percent, Basidiomycotina about 25.2 percent, and Deuteromycotina about 26.8 percent.

Most edible fungi belong to divisions 2 to 5, just listed. Yeasts are single-celled fungi that reproduce asexually by budding or fission, or sexually through ascospore formation. The term “mushroom” refers to those macrofungi (visible to the naked eye) with edible fruiting bodies (sporophores), whereas “toadstool” refers to macrofungi with toxic fruiting bodies; both mushrooms and toadstools are found in more than one of the fungal divisions (Hawksworth et al. 1983; Koivikko and Savolainen 1988).

Historical Background

Fungi have been associated with humans since prehistoric times and must have been collected and eaten along with other plants by hunter-gatherers prior to the development of agriculture (Oakley 1962; Monthoux and Lündstrom-Baudois 1979; Pöder, Peintner, and Pümpel 1992). Although their prehistoric use remains uncertain, they may have been employed as food, in the preparation of beverages, and as medicine.

There is, however, no specific evidence for the use of fungi prior to the Neolithic period, when fungi consumption would have been associated with the drinking of mead (yeast-fermented diluted honey) and yeast-fermented beer or wine, and, some-

what later, the eating of yeast-fermented (leavened) bread.

Mesopotamia

Beer was the preferred fermented drink of the Sumerians of the Late Uruk period dating to the late fourth millennium B.C. R. H. Michel, P. E. McGovern, and V. R. Badler (1992) have noted the similarity of the grooves on a Late Uruk jar at Godin Tepe (in the Zagros Mountains of Iran) with the Sumerian sign for beer, *kas*. The grooves contained a pale yellow residue, which the authors thought was an oxalate salt – oxalate salts are the principal molecules in “beerstone,” a material found on barley beer fermentation containers. The Sumerians were very fond of beer and brewed at least 19 different kinds. They also used beer in some of their medical prescriptions (Majno 1975). Date and grape wine were known in Babylonia by 2900 B.C. (Saggs 1962).

Egypt

Fungi were eaten or drunk, perhaps unwittingly, in Egypt thousands of years ago, in beer and, later, in wine and bread. Yeasts were discovered in a vase of the Late Predynastic period (3650–3300 B.C.) (Saffirio 1972). J. R. Geller (1992) identified a brewery at Hierakonpolis by examining the chemistry of the black residue found in the brewery vats dating from roughly the same time. Similar beer vat sites have been discovered in Egypt from the Amratian (about 3800–3500 B.C.) through the Early Dynastic period (about 3100–2686 B.C.) (Geller 1989).

A yeast, resembling modern *Saccharomyces* spp. and named *Saccharomyces winlocki*, was found in an undisturbed Theban tomb of the Eleventh Dynasty (2135–2000 B.C.) (Gruss 1928). *S. winlocki* was also found in an amphora containing beer in the tomb of



Field mushroom

Queen Meryet-Amun of the Eighteenth Dynasty (1570–1305 B.C.) at Thebes (Winlock 1973).

The bag press, used to press grapes in the manufacture of wine, is shown on the walls of Egyptian tombs from around 3000 B.C. (Forbes 1967). Wine was drunk by upper-class Egyptians and beer was the beverage of the lower classes. Leavened bread was also a common food, as is illustrated by paintings in the tomb of Ramses III (Goody 1982).

Sudan

According to H. A. Dirar (1993), date wine, beer, bread, and cake may have been made in the Sudanese Kingdom of Yam (2800–2200 B.C.). Two men drinking beer or wine through a bamboo straw are shown in a drawing at Mussawarat es Sufra, a site dating to Meroitic times (from between 1500 and 690 B.C. to A.D. 323). Strabo (7 B.C.) mentioned that the Meroites (Ethiopians) knew how to brew the sorghum beer that is called *merissa* today. Wine, which dated from between 1570 and 1080 B.C., was introduced into the Sudan by Egyptian colonists of the New Kingdom (Dirar 1993).

China

In Chinese folklore, Shen Nung, the “Divine Ploughman,” a mythical ruler, taught the people how to use plant medicines and, presumably, taught them about fungi as well. Y. C. Wang (1985) has suggested that the Chinese knew about fungi some 6,000 to 7,000 years ago but offered no specific evidence for their use as food. K. Sakaguchi (1972), who wrote that mold fermentation was traceable to about 1000 B.C. in China, has been supported by T. Yokotsuka (1985).

B. Liu (1958) claimed that in China, mushrooms were first eaten in the Chou Dynasty about 900 B.C. Lui Shi Chuen Zhou, in his *Spring and Autumn of Lui's Family*, recorded the eating of *ling gi* (*Ganoderma* sp.) about 300 B.C. (Liu 1991). The *Book of Songs*, printed in the Han Dynasty (26 B.C. to A.D. 220), mentioned over 200 useful plants, including a number of common edible mushrooms (Wang 1985). Similarly, the *Book of Rites*, written about A.D. 300, mentions several edible fungi (Wang 1985).

Auricularia auricula and *Auricularia polytrica* (“wood ear”) were described by Hsiang Liu (about 300–200 B.C.) and by Hung Wing T'ao (between A.D. 452 and 536). The *T'ang Pen Ts'ao* of the Tang Dynasty (seventh century) described five kinds of *mu-erb*, which grew on various trees. The cultivation of the mushrooms (*Auricularia* spp.) was begun in the Tang Dynasty (A.D. 618–907) (Chang and Miles 1987), although they were probably eaten at least 1,400 years ago (Lou 1982).

The mushroom *Lentinus edodes* (*shiitake*) was recognized as the “elixir of life” by a famous Chinese physician, Wu Shui, during the Ming Dynasty

(1368–1644). This was testimony to a primitive form of *shiitake* cultivation, called *boang-ko*, that had been developed about 800 years ago (Ito 1978). According to R. Singer (1961), Japanese Emperor Chuai was offered the singular mushroom by the natives of Kyushu in A.D. 199, from which we may infer that they were gathered for consumption much earlier in China.

The Chinese method of producing alcoholic beverages required that *Rhizopus*, *Mucor*, or *Saccharomyces* spp. be grown spontaneously on compact bricks of wheat flour or other materials called *kyokushi*. This process was said to have been introduced into Japan at the beginning of the fifth century A.D. and, hence, must have been used appreciably earlier in China (Kodama and Yoshizawa 1977).

A pigment from a fungus was mentioned in *Jib Yang Pen Chaio* (Daily Herb), written by Jui Wu in A.D. 1329. The organism producing the pigment was the yeast *Monascus* sp. (Wang 1985), which grows on rice and has been widely used in the Orient. It is the source of a red pigment employed to color such things as wine and soybean cheese (Lin and Lizuka 1982).

The Mongolian conquests introduced another source of fungal foods – cheese – to the Chinese people, who generally shunned dairy products. Su Hui, royal dietician during the reign of Wen Zong (Tuq. Temur) from A.D. 1328 to 1332, wrote *Yenishan Zhengyao* (The True Principles of Eating and Drinking). In it he included dairy products such as fermented mare's milk, butter, and two cheeses (Sabban 1986). Another dietician of the Yuan Dynasty, Jia Ming (A.D. 1268–1374), also discussed the use of cheese over vegetables or pasta and mentioned fungi as food (Sabban 1986).

As previously hinted, cultivation (rather than gathering from the wild) of mushrooms for human food on a large scale may first have begun in China as early as the Han Dynasty (206 B.C. to A.D. 9). In the first century A.D., Wang Chung's *Lun Heng* stated that the cultivation of *chih* (fungus) was as easy as the cultivation of beans. In 1313, procedures for mushroom cultivation were described in Wong Ching's *Book of Agriculture* (Chao Ken 1980).

Fermented protein foods have an ancient history in China (Yokotsuka 1985). According to the *Shu-Ching*, written about 3,000 years ago, *chu* (yeast or fungus) was essential for the manufacture of alcoholic beverages from wheat, barley, millet, and rice as early as the Chou Dynasty, 1121–256 B.C. By the Han Dynasty, *chu* was made in the form of a cake called *ping-chu*. A sixth-century text on agricultural technology, *Chi-Min-Yao Shu*, detailed the preparation of several kinds of *chu* and other fermented foods such as *chiang* (fermented animal, bird, or fish flesh with millet). *Chu* was a common flavoring in the China of the Chou Dynasty (1121–256 B.C.), and *chiang* was mentioned in the *Analects of Confucius*, written some 600 years after that period. S. Yoshida (1985) wrote that fermented

soybeans originated in China in the Han Dynasty and were known as *shi*.

Greece and Rome

That the ancient Greeks used fungi as food seems clear, because accidental mushroom poisoning was mentioned in the fifth century B.C. by both Euripides and Hippocrates (Buller 1914–16). Theophrastus (d. 287 B.C.) apparently knew and named truffles, puffballs, and fungi (Sharples and Minter 1983).

The Romans enjoyed *boleti* (the *Agaricus* of today) and even had special vessels, called *boletari*, to cook the fungi (Grieve 1925). Presumably, a dish of *boleti* concealed the poisonous mushrooms that Agrippina administered to her husband, the Emperor Claudius, so that her son, Nero, could become emperor of Rome (Grieve 1925).

According to J. André (1985), the Romans ate *Amanita caesarea*, *Boletus purpureus*, and *Boletus suillus*, as well as truffles, puffballs, and morels (Rolfe and Rolfe 1925). Fungi must have been prized by wealthy Romans, for they are mentioned as special delicacies by Horace (65–8 B.C.), Ovid (43 B.C. to A.D. 19), Pliny (A.D. 46–120), Cicero (A.D. 106–143), and Plutarch (A.D. 46–120) (Rolfe and Rolfe 1925; Watling and Seaward 1976). The oldest cookbook presently known was written by Caelius Apicius in the third century A.D. and includes several recipes for cooking fungi (Findlay 1982).

Japan

The earliest reference to mushrooms in Japanese texts is in the *Nihongi* (Book of Chronicles), completed in A.D. 720, which recorded that mushrooms were presented to the Emperor Ojin in A.D. 288 by the local chieftains in Yamato (Wasson 1975). But according to Singer (1961), the earliest consumption of fungi in Japan was in A.D. 199, when the Emperor Chuai was offered *shiitake* by the natives of Kyushu. Mushrooms are rarely mentioned in the early poetry of Japan, but *Manyoshu*, the first anthology of poetry (compiled in the latter half of the eighth century), refers to the pine mushroom, and the *Shui Wakashu* (from about A.D. 1008) mentions it twice. In the *Bunrui Haiku Zenshu*, written by Masaoka Shiki sometime around the beginning of the sixteenth century, there were 250 verses about mushrooms and mushroom gathering (Blyth 1973).

Mexico

The Spanish conquerors of Mexico reported in the sixteenth century that the Aztecs used a mushroom called *teonanacatl* (“god’s flesh”), and sacred mushrooms were pictured in the few Mayan manuscripts that survived the Spanish destruction of “idols and pagan writings.” The Mayan *Codex Badianus*, written in 1552 by Martin de la Cruz, an Indian herbalist, mentioned the use of *teonanacatl* for painful ailments.

The *Codex Magliabecchi* (c. 1565) includes an illustration depicting an Aztec eating mushrooms, and Franciscan friar Bernardino de Sahagun (1499–1590) discussed, in his *General History of the Things of New Spain*, the use of *teonanacatl* to induce hallucinations (Guerra 1967). The Aztecs were familiar enough with fungi to give them names: *nanacatl* (mushroom), *teonanacatl* (sacred mushroom), and *quaubtlanamacatl* (wild mushroom). Indeed, the Mazatecs of Oaxaca and the Chinantecs of Mexico still use hallucinogenic mushrooms for divination, medical diagnosis, and religious purposes (Singer 1978).

The Near East

Al-Biruni, an Arab physician of about 1,000 years ago, described the eating of several fungi, including truffles (Said, Elahie, and Hamarneh 1973). *Terfazia urenaria* is the truffle of classical antiquity, and it is prized in the Islamic countries of North Africa and the Near East as *terfaz*. The best truffles were reputed to come from the areas of Damascus in Syria and Olympus in Greece (Maciarello and Tucker 1994).

Europe

Truffles were already a part of Roman cuisine by the first century A.D., when the Roman poet and satirist Decimus Junius Juvenalis wrote: “[T]he Truffles will be handed round if it is Spring, and if the longed-for thunders have produced the precious dainties.” At that time, fungi were thought to originate when lightning struck the earth during thunderstorms. Truffles were a part of French cuisine by the time of the Renaissance and were exported to England by the beginning of the eighteenth century (Maciarello and Tucker 1994).

In France, mushrooms were cultivated on manure from horse stables during the reign of Louis XIV (Tounefort 1707), and F. Abercrombie (1779) described an English method of composting such manure for the growth of mushrooms by stacking it, a method still in use today. Mushrooms are still highly prized as food in Europe. Many wild fungi are gathered and eaten, and many more are cultivated or imported (Mau, Beelman, and Ziegler 1994).

Fungi Eaten Now and in the Past by Humans

Fungi have been a prized food of peoples past and present around the world. Many examples of these fungi are listed in Table II.C.7.1, which is meant to be indicative rather than exhaustive.

Fungi are mostly eaten cooked, although some ethnic groups and individuals eat them raw. Today, the people of Asia appear to be the most eclectic consumers of fungi. The Chinese eat perhaps as

many as 700 wild and domesticated species. The Japanese use well over 80 species (Imai 1938); the people of India may consume more than 50 species; and the French, not to be outdone, enjoy well over 200 species from one area alone – that of Haute-Savoie (Ramain 1981). Similarly, North Americans eat more than 200 wild and cultivated fungal species (Lincoff 1984).

The reader should be aware that many mushroom genera include both edible and toxic species, and that some mushroom varieties can be edible, whereas others of the same species are not. In the case of some mushrooms, boiling in water before eating will remove toxic or unpleasant secondary metabolites.

Relatively barren areas of the Near East, including parts of Africa and Asia, support thriving populations of truffles, genus *Tirmania*, which are eaten from Morocco and Egypt in North Africa to Israel, Saudi Arabia, and Iraq (Said et al. 1973; Alsheikh, Trappe, and Trappe 1983). Truffles called *fuga* are prized in Kuwait and eaten with rice and meat (Dickson 1971). In some areas of the Arabian Gulf, the truffle crop may be appropriated by the local royal families (Alsheikh et al. 1983).

Today, edible fungi are cultivated or collected in the wild in huge numbers and shipped by air from the source country to consumer countries around the world; fungi may also be canned or dried for long-term storage and later consumption.

Table II.C.7.1. *Fungi eaten by humans around the world now and in the past*

Species by country	People	Local name	Reference
Central Africa			
<i>Boletus</i> sp.			Schnell 1957
Congo			
<i>Auricularia polytricha</i>			"
<i>Boletus sudanicus</i>			"
<i>Cantharellus aurantiaca</i>			"
<i>Clitocybe castanea</i>			"
<i>Lentinus</i> sp.			"
<i>Lepiota</i> sp.			"
<i>Russula</i> sp.			"
<i>Schulzerea</i> sp.			"
Equatorial Africa			
<i>Leucocoprinus molybdites</i>			"
<i>Volvaria diplasia</i>			"
<i>Volvaria esculenta</i>			"
Ivory Coast			
<i>Hygrophoropsis mangeloti</i>			"
Kenya			
Mushrooms		<i>mbeere</i>	Scudder 1971
Libya			
<i>Terfazia boudieri</i>		<i>chatin</i>	Ahmed, Mohamed, and Hami 1981
Madagascar			
<i>Leucocoprinus molbdites</i>			Schnell 1957
Malawi			
<i>Amanita bingensis</i>	Yao	<i>nakajongolo</i>	Morris 1987
<i>Amanita hemibapha</i>	Yao	<i>katelela</i>	"
<i>Amanita zambiana</i>	Yao	<i>utenga</i>	"
<i>Auricularia auricula</i>		<i>matwe</i>	"
<i>Cantharellus congolensis</i>	Chichewa	<i>riakambuzi</i>	"
<i>Cantharellus longisporus</i>	Yao	<i>makungula</i>	"
<i>Cantharellus tenuis</i>	Chichewa	<i>ngundasuku</i>	"
<i>Clavaria albiramea</i>	Yao	<i>nakambi</i>	"
<i>Lentinus cladopus</i>	Yao	<i>nakatasi</i>	"
<i>Lentinus squarrosus</i>	Yao		"
<i>Psathyrella atroumbonata</i>	Yao	<i>ujonjo^a</i>	"
<i>Psathyrella candolleana</i>	Chichewa	<i>nyonzive^a</i>	"
<i>Russula atropurpurea</i>			"

(continued)

Table II.C.7.1. (Continued)

Species by country	People	Local name	Reference
Malawi			
<i>Russula delicata</i>	Chichewa	<i>kamatbova</i>	Morris 1987
<i>Russula lepida</i>	Chichewa	<i>kafidi^a</i>	"
<i>Russula schizoderma</i>	Yao	<i>usuinda^a</i>	"
<i>Strobilomyces constatispora</i>	Chichewa	<i>chipindi^a</i>	Pegler and Pearce 1980
<i>Termitomyces clypeatus</i>	Chichewa	<i>nyonzwe^a</i>	"
<i>Termitomyces titanicus</i>			"
North Africa			
<i>Terfazia</i> spp.			Alsheikh, Trappe, and Trappe 1983
<i>Tirmanea nivea</i>			"
<i>Tirmanea pinoyi</i>			"
South Africa			
<i>Terfazia</i> sp.	San Bushmen		Lee 1979
West Africa			
<i>Leucoprenus molybdites</i>			Schnell 1957
Zambia			
<i>Amanita zambiana</i>			Pearce 1981
<i>Cantharellus densifolius</i>			Morris 1987
<i>Cantharellus longisporus</i>			"
<i>Lactarius vellereus</i>			"
<i>Lactarius</i> spp.			Pearce 1981
<i>Lentinus cladopus</i>			"
<i>Macrolepiota</i> spp.			"
<i>Russula</i> spp.			"
<i>Schizophyllum commune</i>			"
<i>Termitomyces</i> spp.			"
<i>Termitomyces titanicus</i>	Chewa		"
<i>Termitomyces clypeatus</i>	Bebba		"
Zimbabwe			
<i>Volvaria volvacea</i>			Irvine 1952
India			
<i>Agaricus arvensis</i>			Purkayastha 1978
<i>Agaricus basianilosus</i>			"
<i>Agaricus bisporus</i>			"
<i>Agaricus campestris</i>			Bose and Bose 1940
<i>Auricularia delicata</i>			Verma and Singh 1981
<i>Bovista crocatus</i>			Purkayastha 1978
<i>Bovista gigantea</i>			Bose and Bose 1940
<i>Calocybe indica</i>			Purkayastha 1978
<i>Cantharellus aurantiacus</i>			"
<i>Cantharellus cibarius</i>			"
<i>Cantharellus minor</i>			"
<i>Calvatia cyathiformis</i>			"
<i>Clavaria aurea</i>			Verma and Singh 1981
<i>Clitocybe</i> sp.			"
<i>Collybia albuminosa</i>			Purkayastha 1978
<i>Coprinus atramentarius</i>			Kaul 1981
<i>Coprinus comatus</i>			"
<i>Coprinus micaceus</i>			"
<i>Elvela crispa</i>			Purkayastha 1978
<i>Elvela metra</i>			"
<i>Enteloma macrocarpum</i>			Bose and Bose 1940
<i>Enteloma microcarpum</i>			"
<i>Geaster</i> sp.			"
<i>Geopora arenicola</i>			Kaul 1981
<i>Lactarius</i> sp.			Verma and Singh 1981
<i>Lentinus edodes</i>			"
<i>Lentinus subnudus</i>			Purkayastha 1978
<i>Lepiota albuminosa</i>			"

Table II.C.7.1. (Continued)

Species by country	People	Local name	Reference
India			
<i>Lepiota mastoidea</i>			Purkayashtha 1978
<i>Lycoperdon pusillum</i>			"
<i>Lycoperdon pyriformis</i>			Verma and Singh 1981
<i>Macrolepiota mastoidea</i>			Purkayastha 1978
<i>Macrolepiota procera</i>			"
<i>Macrolepiota rachodes</i>			"
<i>Morcbella angusticeps</i>			Kaul 1981
<i>Morcbella conica</i>			"
<i>Morcbella deliciosa</i>			"
<i>Morcbella hybrida</i>			"
<i>Pleurotus flabellatus</i>			"
<i>Pleurotus fossulatus</i>			"
<i>Pleurotus membranaceus</i>			"
<i>Pleurotus ostreatus</i>			Bose and Bose 1940
<i>Pleurotus salignus</i>			Kaul and Kachroo 1974
<i>Russula</i> sp.			Verma and Singh 1981
<i>Schizophyllum commune</i>			"
<i>Scleroderma</i> sp.			"
<i>Termitomyces albuminosa</i>			Purkayastha 1978
<i>Termitomyces eurbizus</i>			"
<i>Termitomyces microcarpus</i>			"
<i>Tricholoma giganteum</i>			Verma and Singh 1981
<i>Verpa bohemica</i>			Kaul 1981
<i>Volvariella diplasia</i>			Purkayastha 1978
<i>Volvariella terastius</i>			"
<i>Volvariella volvacea</i>			"
Indonesia			
<i>Lentinus edodes</i>			Hiepko and Schultze-Motel 1981
<i>Lentinus novopommeranus</i>			"
<i>Oudemansiella apalosorca</i>			"
Japan			
<i>Agaricus arvensis</i>			Imai 1938
<i>Agaricus campestris</i>			"
<i>Agaricus hortensis</i>			"
<i>Agaricus placomyces</i>			"
<i>Agaricus silvaticus</i>			"
<i>Agaricus silvicol</i>			"
<i>Agaricus subrufescens</i>			"
<i>Armillaria caligata</i>			"
<i>Armillaria Matsutake</i>			"
<i>Armillaria mellea</i>			"
<i>Armillaria ventricosa</i>			"
<i>Cantharellus cibarius</i>			"
<i>Cantharellus floccosus</i>			"
<i>Clitocybe extenuata</i>			"
<i>Clitocybe nebularis</i>			"
<i>Clitopilus caespitosus</i>			"
<i>Collybia butyracea</i>			"
<i>Collybia nameko</i>			Ito 1917
<i>Collybia velutipes</i>			"
<i>Cortinarius elatus</i>			Imai 1938
<i>Cortinarius fulgens</i>			"
<i>Cortinarius latus</i>			"
<i>Cortinarius multififormis</i>			"
<i>Cortinellus edodes</i>			"
<i>Cortinellus scalpuratus</i>			"
<i>Cortinellus vaccin</i>			"
<i>Entoloma chypeatum</i>			"
<i>Gomphidius rutilis</i>			"
<i>Gymopilus lentus</i>			"
<i>Gymopilus lubricus</i>			"

(continued)

Table II.C.7.1. (Continued)

Species by country	People	Local name	Reference
Japan			
<i>Gomphidius rutilis</i>			Imai 1938
<i>Gymopilus lentus</i>			"
<i>Gymopilus lubricus</i>			"
<i>Hebeloma mesophaeum</i>			"
<i>Hygrophorus chrysodon</i>			"
<i>Hygrophorus erubescens</i>			"
<i>Hygrophorus pudorinus</i>			"
<i>Hypholoma lateritium</i>			"
<i>Lactarius akabatsu</i>			Tanaka 1890
<i>Lactarius deliciosus</i>			Imai 1938
<i>Lactarius flavidulus</i>			"
<i>Lactarius hatsudake</i>			Tanaka 1890
<i>Lactarius luteolus</i>			Imai 1938
<i>Lactarius piperatus</i>			"
<i>Lactarius sanguifluus</i>			"
<i>Lactarius torminosus</i>			"
<i>Lactarius vellereus</i>			"
<i>Lactarius volemus</i>			"
<i>Lepiota naucina</i>			"
<i>Marasmius oreades</i>			"
<i>Pboliota adiposa</i>			"
<i>Pboliota erebia</i>			"
<i>Pboliota Nameko</i>			"
<i>Pboliota praecox</i>			"
<i>Pboliota squarrosa</i>			"
<i>Pboliota squarrosoides</i>			"
<i>Pboliota terrestris</i>			"
<i>Pboliota togularis</i>			"
<i>Pboliota vablui</i>			"
<i>Pleurotus cornucopiae</i>			"
<i>Pleurotus ostreatus</i>			"
<i>Pleurotus porrigens</i>			"
<i>Pleurotus seriotinus</i>			"
<i>Russula aurata</i>			"
<i>Russula cyanoxantha</i>			"
<i>Russula delica</i>			"
<i>Russula integra</i>			"
<i>Russula lactea</i>			"
<i>Russula virescens</i>			"
<i>Tricholoma albobrunneum</i>			"
<i>Tricholoma cartilagineum</i>			"
<i>Tricholoma conglobatum</i>			"
<i>Tricholoma equestre</i>			"
<i>Tricholoma gambesum</i>			"
<i>Tricholoma humosum</i>			"
<i>Tricholoma nudum</i>			"
<i>Tricholoma personatum</i>			"
<i>Tricholoma pessundatum</i>			"
<i>Tricholoma sejunctum</i>			"
<i>Cortinellus Berkeleyana</i>	Ainu		Yokayama 1975
<i>Grifola frondosa</i>			"
<i>Laetiporus sulphureus</i>			"
<i>Lentinus edodes</i>			"
<i>Panellus serotinus</i>			"
Malaysia			
<i>Amanita manginiana</i>			Burkill 1935
<i>Amanita virginea</i>			"
<i>Psalliota campestris</i>			"
Philippines			
<i>Agaricus argyrostectus</i>			Reinking 1921
<i>Agaricus boltoni</i>			"
<i>Agaricus luzonensis</i>			"
<i>Agaricus manilensis</i>			"

Table II.C.7.1. (Continued)

Species by country	People	Local name	Reference
Philippines			
<i>Agaricus merrillii</i>			Reinking 1921
<i>Agaricus perfuscus</i>			"
<i>Auricularia auricula-judea</i>			"
<i>Auricularia brasiliensis</i>			"
<i>Auricularia cornea</i>			"
<i>Auricularia tenuis</i>			"
<i>Boletus</i> spp.			"
<i>Collybia albuminosa</i>			"
<i>Coprinus ater</i>			"
<i>Coprinus Bagobos</i>			"
<i>Coprinus bryanti</i>			"
<i>Coprinus concolor</i>			"
<i>Coprinus confertus</i>			"
<i>Coprinus deliquescens</i>			"
<i>Coprinus fimbriatus</i>			"
<i>Coprinus flos-lactus</i>			"
<i>Coprinus friesii</i>			"
<i>Coprinus nebulosus</i>			"
<i>Coprinus ornatus</i>			"
<i>Coprinus plicatilis</i>			"
<i>Coprinus pseudo-plicatus</i>			"
<i>Coprinus volutus</i>			"
<i>Cortinarius</i> spp.			"
<i>Daedalea</i> spp.			Yen and Gutierrez 1976
<i>Lentinus esculenta</i>			Reinking 1921
<i>Lentinus pruinosa</i>			"
<i>Lepiota candida</i>			"
<i>Lepiota cepaestipes</i>			"
<i>Lepiota chlorospora</i>			"
<i>Lepiota elata</i>			"
<i>Lepiota fusco-squamea</i>			"
<i>Lepiota revelata</i>			"
<i>Lepiota sulphopenita</i>			"
<i>Lycoperdon cepiformis</i>			"
<i>Lycoperdon furfuraceum</i>			"
<i>Lycoperdon lilacinum</i>			"
<i>Lycoperdon polymorphum</i>			"
<i>Lycoperdon pyriforme</i>			"
<i>Marasmius</i> spp.			"
<i>Panaeolus panaiense</i>			"
<i>Panaeolus papilionaceus</i>			"
<i>Panaeolus pseudopapilionaceus</i>			"
<i>Panaeolus veluticeps</i>			"
<i>Pleurotus applicatus</i>			"
<i>Pleurotus noctilucens</i>			"
<i>Pleurotus ostreatus</i>			"
<i>Pleurotus striatulus</i>			"
<i>Tricholoma tenuis</i>			"
Singapore			
<i>Amanita virginea</i>			Burkill 1935
Thailand			
<i>Termitomyces fuliginosus</i>			Bels and Pataragetvit 1982
<i>Termitomyces globulus</i>			"
<i>Termitomyces mammiformis</i>			"
<i>Termitomyces microcarpus</i>			"
Vietnam			
<i>Amanita manginiana</i>			Burkill 1935
Australia			
<i>Agaricus campestris</i>			Irving 1957
<i>Elderia arenivaga</i>	Aborigine		Cribb and Cribb 1975
<i>Mylitta australis</i>	Tasmania		Irving 1957

(continued)

Table II.C.7.1. (Continued)

Species by country	People	Local name	Reference
New Caledonia			
<i>Agaricus edulis</i>			Barrau 1962
<i>Hydnum caputmedusae</i>			"
Papua New Guinea			
<i>Amanita bemibapba</i>			Shaw 1984
<i>Boletus</i> spp.			"
<i>Clitocybe</i> spp.			"
<i>Cortinarius</i> spp.			"
<i>Heimiella rubrapuncta</i>			"
<i>Lentinus edodes</i>			"
<i>Lentinus tuber-regium</i>			"
<i>Lepiota</i> spp.			"
<i>Pleurotus djamor</i>			"
<i>Polyporus frondosus</i>			"
<i>Polyporus sulphureus</i>			"
<i>Psalliota</i> spp.			"
<i>Tremella fuciformis</i>			"
New Zealand			
<i>Agaricus adiposus</i>	Maori		Colenso 1881
<i>Hirneola auricula-judae</i>	Maori		"
<i>Inocybe cibarium</i>	Maori		"
<i>Lycoperdon fontanesei</i>	Maori		"
<i>Lycoperdon giganteum</i>	Maori		"
Europe (general)			
<i>Boletus aereus</i>			Singer 1961
<i>Boletus edulis</i>			"
<i>Cantbarellus cibarius</i>			Johnson 1862
<i>Morcbella esculenta</i>			"
<i>Polyporus squamosus</i>			"
Austria			
<i>Armiliariella mellea</i>			Singer 1961
England			
<i>Agaricus campestris</i>			Johnson 1862
<i>Agaricus deliciosus</i>			"
<i>Agaricus oreades</i>			"
<i>Agaricus personatus</i>			"
France			
<i>Agaricus prunulus</i>			"
<i>Boletus edulis</i>			"
<i>Tuber cibarium</i>			"
Germany			
<i>Cantbarellus cibarius</i>			Singer 1961
Italy			
<i>Agaricus prunulus</i>			Johnson 1862
<i>Tuber cibarium</i>			"
Poland			
<i>Boletus edulis</i>			Usher 1974
Russia			
<i>Boletus versipellis</i>			"
<i>Lactarius scrobiculatus</i>			Singer 1961
Spain			
<i>Lactarius deliciosus</i>			"
<i>Lactarius sanguifluus</i>			"
Switzerland			
<i>Cantbarellus cibarius</i>			"

Table II.C.7.1. (Continued)

Species by country	People	Local name	Reference
Mexico			
<i>Saccharomyces</i> sp.		<i>chicha bruja</i>	Singer 1961
<i>Boletus fragrans</i>	Chinantec		Lipp 1991
<i>Schizophyllum commune</i>	Chinantec		"
<i>Ustilago maydis</i>	Chinantec		"
<i>Coriolus</i> sp.	Maya		Alcorn 1984
<i>Lentinus</i> cf. <i>lepideus</i>	Huastec, Maya		"
<i>Schizophyllum commune</i>	Huastec, Maya		"
<i>Ustilago maydis</i>	Huastec, Maya		"
<i>Cantbarellus cibarius</i>	Mixe		Lipp 1991
<i>Pleurotus</i> sp.	Mixe		"
<i>Ramaria</i> sp.	Mixe		"
<i>Schizophyllum commune</i>	Mixe		"
<i>Ustilago maydis</i>	Nahua		Martinez et al. 1983
<i>Ustilago maydis</i>	Purepeche		Mapes, Guzman, and Cabellero 1981
<i>Ustilago maydis</i>	Totonacs		Martinez et al. 1983
<i>Ustilago maydis</i>	Teenek		Alcorn 1984
<i>Lentinus lepideus</i>	Teenek		"
<i>Schizophyllum commune</i>	Teenek		"
North America			
<i>Agaricus campestris</i>	Yaki Indians		Mead 1972
<i>Agaricus campestris</i>	Iroquois, Straits Salish		Kuhnlein and Turner 1991
<i>Armillaria mellea</i>	Flathead Indians		Turner 1978
<i>Boletus</i> sp.	Calpella Indians		Chestnut 1902
<i>Bovista plumbea</i>	Omaha Indians		Yanovsky 1936
<i>Calvatia gigantea</i>	Iroquois, Upriver Halkomelem		Kuhnlein and Turner 1991
<i>Calvatia gigantea</i>	Omaha Indians		Yanovsky 1936
<i>Cantbarellus cibarius</i>	Nlaka'pamux		Kuhnlein and Turner 1991
<i>Collybia</i> spp.	Flathead Indians		Hart 1979
<i>Fistulina hepatica</i>	Cherokee		Hamil and Chiltoskey 1975
<i>Ganoderma applanatum</i>	Halkomelem		Kuhnlein and Turner 1991
<i>Inonotus obliquus</i>	Woods Cree		"
<i>Lycoperdon giganteum</i>	Iroquois		Yanovsky 1936
<i>Lycoperdon</i> sp.	White Mtn. Apache		Reagan 1929
<i>Lycoperdon</i> sp.	Flathead Indians		Hart 1979
<i>Morchella</i> sp.	Iroquois		Arnason, Hebela, and Johns 1981
<i>Morchella</i> sp.	Lillooet, Halkomelem		Kuhnlein and Turner 1991
<i>Pleurotus ostreatus</i>	Interior Salish		"
<i>Polyporus frondosus</i>	Iroquois		"
<i>Polyporus pinicola</i>	Iroquois		"
<i>Polyporus sulphureum</i>	Iroquois		"
<i>Polystictus versicolor</i>	Dakota Indians		Yanovsky 1936
<i>Russula</i> spp.	Flathead		Hart 1979
<i>Tremelodon</i> spp.	Straits Salish		Kuhnlein and Turner 1991
<i>Tricholoma</i> spp.	Nlaka'pamax		"
<i>Tricholoma magnivelare</i>	Interior Salish		"
<i>Tricholoma populinum</i>	Interior Salish		"
Argentina			
<i>Agaricus campeanus</i>	Onas		Stuart 1977
<i>Cyttaria darwinia</i>	Onas		"
<i>Cyttaria bariotii</i>	Onas		"
<i>Cyttaria bookeri</i>	Onas		"
<i>Fistulina</i> sp.	Onas		"
<i>Polyporus eucalyptorum</i>	Onas		"
<i>Pycnoporus sanguinoreus</i>	Lengua Maskoy		Arenas 1981
Brazil			
<i>Coriolus zonatus</i>	Yanomamo		Prance 1983
<i>Favolus brunneolus</i>	Yanomamo		"
<i>Favolus striatulus</i>	Yanomamo		"
<i>Favolus waikassamo</i>	Yanomamo		"
<i>Gymnopilus hispidellus</i>	Yanomamo		"
<i>Hexagona subcaperata</i>	Yanomamo		"

(continued)

Table II.C.7.1. (Continued)

Species by country	People	Local name	Reference
Brazil			
<i>Hydnopolyporus palmatus</i>	Yanomamo		Prance 1983
<i>Lactocollybia aequatorialis</i>	Yanomamo		"
<i>Lentinus crinitis</i>	Yanomamo		"
<i>Lentinus glabratus</i>	Yanomamo		"
<i>Lentinus velutinus</i>	Yanomamo		"
<i>Neoclitocybe bisseida</i>	Yanomamo		Prance 1972
<i>Panus rudis</i>	Yanomamo		"
<i>Pboliota bicolor</i>	Yanomamo		"
<i>Pleurotus concavus</i>	Yanomamo		"
<i>Polyporus</i> sp.	Yanomamo		"
<i>Polyporus aquasus</i>	Yanomamo		"
<i>Polyporus dermoporus</i>	Yanomamo		"
<i>Polyporus stipitarius</i>	Yanomamo		"
Colombia			
<i>Fomes ignarius</i>			Montes 1961
Ecuador			
<i>Auricularia fuscusuccinea</i>	Waorani		Davis and Yost 1983
Peru			
<i>Auricularia nigrescens</i>	Aymara		Herrera 1934
<i>Daedalea repanda</i>			"
<i>Galera</i> sp.			"
<i>Henzites striata</i>			"
<i>Hirneola polytricha</i>			"
<i>Lycoperdon</i> sp.			"
<i>Schizophyllum alneum</i>			"
<i>Schizophyllum commune</i>			"
<i>Ustilago maydis</i>			Gade 1975

^a*ndiwa* (Chichewa) or *mboga* (Yao) is a relish or side dish of mushrooms fried in oil with onions, tomatoes, and ground nut flour.

Gross Chemical Composition of Fungi

The gross chemistry of edible fungi (Table II.C.7.2) varies with the stage of the life cycle in which they are eaten; for example, the mycelium of *Agaricus campestris*, a common white mushroom, contains 49 percent protein (Humfeld 1948), whereas the sporophore of the same species is 36 percent protein (McConnell and Esselen 1947).

Even the stage in the life cycle of the sporophore may significantly affect the gross chemistry of the fungus (Table II.C.7.3). The sporophore is the fungal part usually eaten, although the mycelium predominates in fermented foods (Purkayastha and Chandra 1976).

Most of the biomass of fungi is water, although there are wide variations in the amount of water in different species (Table II.C.7.2). The dry biomass is mainly carbohydrates, followed by proteins, lipids, and ash, in that order; again, there is wide variation in the amounts of the major components (Table II.C.7.2). In general, dried fungi contain 2 to 46 percent protein, 5 to 83 percent carbohydrates, 1 to 26 percent lipids, 1 to 10 percent RNA, 0.15 to 0.3 percent DNA, and 1 to 29 percent ash (Griffin 1981). Fungal strains with unusually high lipid content have been selected for

this trait and are grown under conditions where lipid synthesis is enhanced. These fungi serve as valuable sources of lipids that are required in large quantities for industrial purposes.

The nutritional value of many edible fungi compares well with other common foods. In essential amino acid content, where meat rates 100 and milk 99, mushrooms are rated at 98. Measuring by amino acid "score," meat scores 100, milk scores 91, and mushrooms score 89, whereas, by nutritional index, meat can score between 59 and 35, soybeans score 31, and mushrooms score 28. Indeed, by any of these criteria, some mushrooms have more nutritional value than all other plants except soybeans; at the same time, however, some edible fungi score much lower by the same criteria (Crisan and Sands 1978). One hundred grams of dried fungal biomass has an energy equivalent of from 268 to 412 kilocalories (Griffin 1981).

Table II.C.7.2 indicates that fungi provide significant amounts of protein. There is, however, some question as to how much of fungal protein is digestible (Crisan and Sands 1978). Fungi also contain sufficient quantities of the essential amino acids required by humans and other animals and have a variety of other nitrogen-containing molecules.

Table II.C.7.2. Gross chemical composition of fungi as a percentage of fungal dry weight

Species	No. of samples	Protein	Total carbohydrate plus fiber	Total lipids	Total nucleic acids	Ash	H ₂ O ^a	Ref.
<i>Agaricus bisporus</i>	5	25-33	65-72	2-3	-	8-12	90	1
<i>Agaricus bisporus</i>		30	66	5	-	17	-	4
<i>Agaricus bisporus</i>		30	-	-	-	-	-	5
<i>Agaricus campestris</i>		33	65	2	-	8	90	1
<i>Agaricus merrilli</i>		34	61	2	-	10	89	1
<i>Agaricus perfuscus</i>		35	63	2	-	11	91	1
<i>Armillariella mellea</i>		28	101	6	-	15	-	4
<i>Auricularia polytrica</i>	2	16	-	-	-	-	-	5
<i>Boletus aestivalis</i>		6	97	6	-	5	89	1
<i>Boletus aestivalis</i>		45	35	6	-	15	-	4
<i>Boletus edulis</i>		30	-	-	-	-	-	5
<i>Candida utilis</i>		30	68	3	-	8	87	1
<i>Cantibarellus cibarius</i>	2	47	27	5	-	9	6	6
<i>Cantibarellus cibarius</i>		22	76	5	-	8	92	1
<i>Clavaria botrytis</i>		18	87	8	-	11	-	4
<i>Clitocybe multiceps</i>		9	82	3	-	6	89	1
<i>Collybia albuminosa</i>		24	69	6	-	12	94	1
<i>Collybia sp.</i>		21	76	4	-	7	89	1
<i>Coprinus atramentarius</i>	2	5	70	4	-	26	95	1
<i>Coprinus comatus</i>		21	66	4	-	19	93	1
<i>Coprinus comatus</i>		25	66	3	-	13	92	1
<i>Coprinus comatus</i>		23	68	5	-	13	-	4
<i>Flammulina velutipes</i>		18	77	2	-	7	89	1
<i>Hirneola affinis</i>		10	77	14	-	11	98	1
<i>Hypoboloma candolleianum</i>		19	77	3	-	14	89	1
<i>Lactarius deliciosus</i>	2	19	68	5	-	6	89	1
<i>Lactarius batsudake</i>		18	79	3	-	8	93	1
<i>Leccinum scabrum</i>		38	58	8	-	14	-	4
<i>Lentinus edodes</i>	2	16	80	6	-	5	91	1
<i>Lentinus edodes</i>		18	76	2	-	7	90	2
<i>Lentinus exilis</i>		11	83	2	-	4	72	1
<i>Lepiota procera</i>		20	76	4	-	7	84	1
<i>Lepista nebularis</i>		40	51	5	-	13	-	4
<i>Lycoperdon lilacinum</i>		46	51	8	-	8	99	1
<i>Macrolepiota procera</i>		32	52	9	-	16	-	4
<i>Marasmius oreades</i>		36	64	8	-	12	-	4
<i>Morchella crassipes^a</i>		2	5	1	-	2	90	7
<i>Morchella esculenta</i>	2	23	72	5	1	10	90	1
<i>Morchella esculenta</i>		3	5	0.3	-	2	90	7
<i>Morchella bortensis^a</i>		3	5	0.3	-	2	90	7
<i>Pholiota nameko</i>		21	73	4	-	8	95	1
<i>Pleurotus eous</i>		18	71	1	-	9	92	2
<i>Pleurotus florida</i>		19	70	2	-	9	92	2
<i>Pleurotus flabellatus</i>		22	69	2	-	11	91	2
<i>Pleurotus limpidus</i>		39	74	9	-	5	93	1
<i>Pleurotus opuntia</i>		9	80	2	-	16	58	1
<i>Pleurotus ostreatus</i>		11	89	2	-	6	74	1
<i>Polyporus sulfureus</i>		14	78	3	-	7	71	1
<i>Russula delica</i>		29	78	9	-	8	-	4
<i>Russula vesca</i>		19	57	6	-	14	-	4
<i>Russula vesca</i>		11	-	-	-	-	-	5
<i>Saccharomyces cerevisiae</i>		37	39	2	8	7	-	9
<i>Suillus granulatus</i>		36	54	7	-	13	-	4
<i>Terfazia boudieri</i>		17	61	6	-	13	78	3
<i>Tricholoma populinum^a</i>		1	4	1	-	1	94	8
<i>Tricholoma portentosum</i>		26	52	4	-	18	-	4
<i>Tuber melanosporum</i>		23	94	2	-	8	77	3
<i>Volvariella diplasia</i>		29	75	3	-	12	90	1
<i>Volvariella esculenta</i>		34	42	21	-	13	91	1
<i>Volvariella volvacea</i>	3	27	68	7	-	9	8	1
<i>Xerocomus subtomentosus</i>		26	51	8	-	14	-	4

^aRefers to samples that were analyzed fresh; all others were analyzed when dry.

References:

1. Crisan and Sands (1978)
2. Bano and Rajarathnam (1982)
3. Ahmed, Mohamed, and Hami (1981)
4. Levai (1989)
5. Chang and Hayes (1978)
6. Sinskey and Batt (1987)
7. Litchfield, Vely, and Overbeck (1963)
8. Turner, Kuhnlein, and Egger (1987)
9. Miller (1968).

Table II.C.7.3. *Variations in the gross chemistry of different stages in the development of the Volvariella volvacea sporophore*

Chemistry (as % dry weight)	Sporophore stage			
	Button	Egg	Elongation	Mature
Moisture	89	89	89	89
Crude fat	1	2	2	4
N-free carbohydrate	43	51	50	40
Crude fiber	6	5	7	13
Crude protein	31	23	21	21
Ash	9	8	9	10
Energy (Kcal/100 g)	281	287	281	254

Source: Li and Chang (1982).

Soluble carbohydrates in fresh fungi range from 37 to 83 percent of the dry weight. In addition, there are fiber carbohydrates that make up from 4 to 32 percent (Griffin 1981).

The lipid content of fungi ranges from 0.2 percent of the cell or tissue dry weight to as high as 56 percent - more specifically, 0.2 to 47 percent for Basidiomycotina, and 2 to 56 percent for Deuteromycotina (Weete 1974; Wassef 1977). Sporophores' contents of

lipids tend to be relatively low, but among them are triglycerides, phospholipids, fatty acids, carotenoids, and steroids, as well as smaller amounts of rarer lipids.

Carotenoids may accumulate in some fungi; in fact, some pigmented fungi have been grown in bulk precisely for carotenoids, which are fed to carp or to chickens to color their eggs and make them more acceptable to the consumer (Klaui 1982). Some of these carotenoids may be converted to vitamin A in humans (Tee 1992).

In addition, some fungi are sufficiently good producers of the B vitamins to make them viable commercial sources of these nutrients. *Saccharomyces* spp., for example, are good sources of B vitamins generally (Umezawa and Kishi 1989), and riboflavin is obtained in goodly amounts from *Ashbya gossypii* fermentations (Kutsal and Ozbas 1989). The water-soluble vitamin content of several fungi are shown in Table II.C.7.4.

Fungal Flavors and Volatiles

The nonvolatile meaty flavors of edible fungi come primarily from the amino acids (glutamic acid is one of the most common), purine bases, nucleotides (such as the shiitake mushroom's guanosine-5'-monophosphate)

Table II.C.7.4. *Vitamin content of edible fungi*

Species	Thiamine (µg/g)	Riboflavin (µg/g)	B ₆ (µg/g)	Nicotinic acid (µg/g)	Pantothenic acid (µg/g)	Vitamin C (µg/100 g)	Ref.
<i>Agaricus bisporus</i>	10-90	40-50	-	430-570	230	27-82	1
<i>Agaricus bisporus</i>	1	5	22	41	-	4	4
<i>Agaricus bisporus</i>	1,100	5,000	-	55,700	-	-	6
<i>Agaricus breitschneideri</i>	560	110	-	5,100	-	-	6
<i>Agaricus campestris</i>	11	50	-	56	23	82	8
<i>Agaricus campestris</i>	5	21	-	191	-	14	2
<i>Auricularia auricula-judea</i>	120	490	-	5,100	-	-	6
<i>Auricularia polytricha</i>	2	9	-	20	-	-	1
<i>Auricularia polytricha</i>	1	2	-	1	-	1	4
<i>Candida utilis</i>	5	44	33	47	37	-	4
<i>Flammulina velutipes</i>	61	52	-	1065	-	46	1
<i>Lactarius batsudake</i>	15	-	-	-	-	-	1
<i>Lentinus edodes</i>	78	49	-	549	-	-	1
<i>Morchella</i> sp.	4	25	6	82	9	-	3
<i>Pleurotus nameko</i>	188	146	-	729	-	-	1
<i>Pleurotus ostreatus</i>	48	47	-	1,087	-	-	1
<i>Saccharomyces</i> (Brewer's)	44	1,210	-	107	-	-	2
<i>Saccharomyces</i> (Brewer's)	156	43	-	379	-	-	5
<i>Torula</i> sp.	53	450	334	4,173	372	-	10
<i>Torula</i> sp.	140	56	-	444	-	-	5
<i>Tricholoma</i> sp.	6	29	-	885	-	52	1
<i>Volvaria esculenta</i>	90	410	-	4,500	-	3	6
<i>Volvariella volvacea</i>	320	1,630	-	47,600	-	-	7
<i>Volvariella volvacea</i>	12	33	-	919	-	20	1

References

1. Crisan and Sands (1978)
2. Adams (1988)
3. Robinson and Davidson (1959)
4. Haytowitz and Matthews (1984)
5. Bano (1978)
6. Tung, Huang, and Li (1961)
7. Li and Chang (1982)
8. Litchfield (1967)
9. Litchfield, Vely, and Overbeck (1963)

(Nakajima et al. 1961), and the products of the enzymatic breakdown of unsaturated fatty acids. The volatile flavors include C8 compounds, such as benzyl alcohol, benzaldehyde, and other compounds (Mau et al. 1994). Many fungi contain monoterpenes, which produce a variety of flavors and odors; *Trametes odorata*, *Phellinus* spp., and *Kluyveromyces lactis*, for example, produce linalool (sweet, rose-like), geraniol (rose-like), nerool (sweet, rose-like) and citronellol (bitter, rose-like).

Fungi also produce flavors and odors that are buttery; nutlike; mushroomlike; coconut-like (*Trichoderma* spp.); peachlike (*Fusarium poae*, *Pityrosporum* spp., *Sporobolomyces odoratus*, *Trichoderma* spp.); flowery and woody (*Lentinus lepideus*); earthy (*Chaetomium globosum*); sweet, aromatic, and vanilla-like (*Bjerkandera adusta*); coconut- and pineapple-like (*Polyporus durus*); sweet and fruity (*Poria aurea*); and passion-fruit-like (*Tyromyces sambucus*) (Kempler 1983; Schreier 1992).

The flavor of truffles, as in other fungi, is partly caused by nonvolatile organic molecules such as those mentioned previously and by over 40 volatile organic molecules. The aroma of white truffles comes mainly from one of the latter, whereas the Perigord (black) truffle's aroma is the result of a combination of molecules. Species of *Russula*, when dried, have an odor that has been attributed to amines (Romagnesi 1967).

Some fungi produce volatile molecules that attract animals – including humans – or emit distinct flavors (Schreier 1992). Truffles, which are among the most valuable of edible fungi, grow underground on plant roots and produce odors that can only be recognized by dogs, pigs, and a few other mammals. Humans cannot smell them, but those with experience can detect the presence of truffles below the ground by cracks that appear in the soil surface over the plant roots.

The major species of truffles are *Tuber melanosporum* – the black truffle or so-called Perigord truffle – found most frequently in France, Italy, and Spain; *Tuber brumale*, also of Europe; *Tuber indicum* of Asia; and *Tuber aestivum* – the summer truffle or “cook's truffle” – which is the most widespread of all the truffles and the only one found in Britain (Pacioni, Bellina-Agostinone, and D'Antonio 1990).

Fungi and Decay

Along with bacteria, fungi hold primary responsibility for the biological process known as decay, in which complex organic molecules are progressively broken down to smaller molecules by microbial enzymes. The decay process destroys toxic molecules and regenerates small molecules used by microbial or plant life. Examples include carbon as carbon dioxide and nitrogen as amino acids, nitrates, or ammonia. Dead plants and animals are broken down to humus and simpler organic molecules that fertilize the soil and increase

its water-holding capacity. These same processes have been used by humans in fermentation and in the production of bacterial and fungal single-cell protein (SCP) from waste or cheap raw materials.

Fungal Fermentation

The process of fermentation or microbial processing of plant or animal foods has served many functions, both now and in the distant past, especially in warm and humid climates where food spoils quickly. Fermentation preserves perishable food at low cost, salvages unusable or waste materials as human or animal food, reduces cooking time and use of fuel, and enhances the nutritional value of food by predigestion into smaller molecules that are more easily assimilated. Sometimes, but not always, fermentation increases the concentration of B vitamins (Goldberg and Thorp 1946) and protein in food (Cravioto et al. 1955; Holter 1988) and destroys toxic, undesirable, or antidiigestive components of raw food. Moreover, fermentation can add positive antibiotic compounds that destroy harmful organisms, and the acids and alcohol produced by fermentation protect against microbial reinfection and improve the appearance, texture, consistency, and flavor of food. In addition, fermented foods often stimulate the appetite (Stanton 1985).

In ancient times, preservation of foods (such as milk, cheese, and meat) and beverages (like beer, mead, and wine) by fermentation made it possible for humans to travel long distances on land or water without the need to stop frequently for water or food. As described by Dirar (1993), over 80 fermented foods and beverages are presently used by the people of the Sudan, including 10 different breads, 10 different porridges, 9 special foods, 13 different beers, 5 different wines, 1 mead, 7 dairy sauces, 4 different meat sauces, 5 different fish sauces, 5 flavors and substitutes of animal sauces, and 10 flavors and substitutes of plant origin.

Today a wide variety of mainly carbohydrate-rich substrates, like cereals, are preserved, but protein-rich legumes and fish can also be processed by fungi. The combination of fungi, yeast, and bacteria is often controlled by antibacterials, fatty acids that act as trypsin inhibitory factors, and phytases, which destroy soybean phytates that bind essential metals (Hesseltine 1985). Table II.C.7.5 lists some of the foods that depend on fungal processing before they may be eaten (Beuchat 1983; Reddy, Person, and Salunkhe 1986).

Fungal fermentation of cereals does not lead to a marked increase in the protein content of the grain, but it does contribute to a significant increase in amino acids, especially those considered essential to humans. There is a decrease in carbohydrates during fungal fermentation, and lipids are hydrolyzed to fatty acids. Fungal fermentation may lead to an increase in B-vitamin content, although B₁₂ will appear only if bacteria are involved in the fermentation.

Table II.C.7.5. *Foods and beverages that require fungal processing*

Product	Substrate	Geographic area	Fungal species
Alcoholic beverages	Cereals, carbohydrates, fruit	Worldwide	<i>Saccharomyces</i> sp.
Ang-kak	Rice	Asia, Syria	<i>Monascus</i>
Banku	Maize, cassava	Ghana	Yeast, bacteria
Bonkrek	Coconut press cake	Indonesia	<i>Rhizopus oligosporus</i>
Burukutu	Sorghum, cassava	Nigeria	<i>Candida</i> spp.
Burung hyphon	Shrimp, fish	Philippines	Yeast
Chee-fan	Soybean whey curd	China	<i>Mucor</i> sp., <i>Aspergillus glaucus</i>
Cheeses			
Brie	Milk curd	France	<i>Penicillium camemberti</i>
Camembert	Milk curd	France	<i>Penicillium camemberti</i>
Gorgonzola	Milk curd	France	<i>Penicillium roqueforti</i>
Roquefort	Milk curd	France	<i>Penicillium roqueforti</i>
Stilton	Milk curd	France	<i>Penicillium roqueforti</i>
Chicha	Maize	Peru	Yeast, bacteria
Colonche	Prickly pears	Mexico	<i>Torulopsis</i> sp.
Dawadawa	Millet	West Africa	Yeast, bacteria
	African locust bean	Nigeria	Yeast, bacteria
Dhoka	Wheat and/or pulses	India	Yeast, bacteria
Dosai	Black gram	India	Yeast, bacteria
Enjera	Teff	Ethiopia	<i>Candida guilliermondii</i>
Fermented manioc	Manioc	Zaire	Yeast, fungi, bacteria
Gari	Cassava	West Africa	<i>Candida</i> spp.
Hamanatto	Soybean and wheat flour	Japan	<i>Aspergillus oryzae</i> and bacteria
Hopper	Rice	Sri Lanka	Yeast, bacteria
Idli	Rice and black gram	India	Yeast, bacteria
Injera	Teff, maize, wheat, barley	Ethiopia	<i>Candida guilliermondii</i>
Jalabies	Wheat flour	India, Nepal, Pakistan	?
Jamin-bang	Maize	Brazil	Yeast and bacteria
Kaanga-kopuwai	Maize	New Zealand	Yeast and bacteria
Kanji	Rice, carrots	India	<i>Hansenula anomala</i>
Kaoling liquor	Sweet fruit or juice	China	<i>Monascus</i> sp.
Oncom	Peanut press cake	Indonesia	<i>Neurospora intermedia</i> , <i>Rhizopus oligosporus</i>
Papadam	Black gram	India	<i>Saccaromyces</i> spp.
Poi	Taro	Hawaii	Fungi and bacteria
Pozol	Maize	Mexico	Fungi and bacteria
Puto	Rice	Philippines	Fungi and bacteria
Sake	Rice, steamed	Japan	<i>Aspergillus oryzae</i> and <i>Saccharomyces cerevisiae</i>
Shoyu	Soy and wheat	Malaysia	<i>Aspergillus oryzae</i> and yeast and bacteria
Sierra rice	Unhusked rice	Ecuador	<i>Aspergillus</i> spp. and bacteria
Sorghum beer	Sorghum, maize	South Africa	Yeast and bacteria
Soy sauce	Soybeans, wheat	Orient	<i>Aspergillus</i> spp., yeast, bacteria
Sufu	Soybean whey curd	China, Taiwan	<i>Actinomucor elegans</i> , <i>Mucor</i> spp.
Tao-si	Soybeans and wheat flour	Philippines	<i>Aspergillus oryzae</i>
Taotjo	Soybeans and wheat or rice	East India	<i>Aspergillus oryzae</i>
Taupe	Cassava or rice	Indonesia	Fungi
Tauco	Soybeans, cereal	West Java	<i>Rhizopus oligosporus</i> , <i>Aspergillus oryzae</i>
Tempeh	Soybeans	Indonesia	<i>Rhizopus</i> spp.
Katsuobushi	Fish	Japan	<i>Aspergillus glaucus</i>
Kecap	Soybeans, wheat	Indonesia	<i>Aspergillus</i> spp., bacteria, yeast
Kenkey	Maize	Africa	Yeast, bacteria, fungi
Khaman	Bengal gram	India	Yeast, bacteria
Kochujang	Hot pepper bean paste	Korea	Yeast, bacteria, fungi
Koji	Rice	China, Japan	<i>Aspergillus oryzae</i>
Lao-cho	Rice	Indonesia, China	<i>Rhizopus</i> spp., other fungi
Meitanza	Soybean cake	China, Taiwan	<i>Actinomucor elegans</i>
Meju	Soybeans	Korea	<i>Rhizopus</i> spp., <i>Aspergillus oryzae</i>
Merissa	Sorghum	Sudan	<i>Saccharomyces</i> spp.
Minchin	Wheat gluten	China	Fungi
Miso	Rice and soybeans, rice and cereals	China, Japan	Fungi
Nan	Wheat	India	Yeast
Nyufu	Fermented dry tofu	East Asia	<i>Actinomucor repens</i>
Ogi	Maize	Nigeria, West Africa	Fungi and bacteria
Torani	Rice	India	<i>Candida</i> spp., <i>Hansenula anomala</i> , other fungi
Waries	Black gram	India	Yeasts
Tuba	Coconut palm sap	Mexico	Yeasts
Kaoliang liquor	Sweet fruit or juice	China	<i>Monascus</i> sp.
Manioc, fermented	Manioc	Zaire	Yeast, bacteria, fungi

Sources: Steinkraus (1983); Jay (1986); Paredes-Lopez and Harry (1988); and Chavan and Kadam (1989).

Soybeans constitute a good example. Normally they contain B vitamins, but neither vitamin B₁₂ nor significant amounts of proteins. When fermented, however, the B vitamins (except for thiamine) increase, and proteins are completely hydrolyzed to amino acids (Murata 1985). Vitamin B₁₂ has been found in all commercial samples of fermented tempe, indicating that bacteria were involved in the fermentation as well (Steinkraus 1985).

In Fiji, carbohydrate-rich crops such as breadfruit (*Artocarpus utilis*), cassava (*Manihot dulcis*), taro (*Colocasia esculenta*), plaintain (*Musa paradisiaca* subsp. *normalis*), banana (*Musa* subsp. *sapientum*), and giant swamp taro (*Alocasia indica*) are preserved for future use by pit-mixed fermentation. This process was probably brought to Tonga during the Lapita period some 2,000 to 3,000 years ago and subsequently spread to Fiji (Aalbersberg, Lovelace Madhaji, and Parekenson 1988).

Single-Cell Protein for Human and Animal Food

Fungi have been employed to produce single-cell protein (SCP) from a variety of waste materials that might otherwise be useless, such as crop straw, bagasse, starchy plant materials, and whey, among others. *Candida* alkane yeasts have been examined for their ability to produce protein-rich biomass and edible calories for pigs and other animals, whereas *Chaetoceros* and *Sporotrichum* spp. have been utilized to enrich the protein content of lignocellulose wastes - like straw - for animal feed. *Rhizopus oligosporus* NRRL 270 has been used to increase the protein content of starchy residues (cassava, potato, and banana), and yeasts have been exploited to produce food and alcohol from whey. Treating manioc with *R. oligosporus* by any of three different fermentation methods has resulted in a marked increase in protein content, seemingly at the expense of the carbohydrate content of the manioc (Ferrante and Fiechter 1983).

Alcoholic Fermentation

In Europe, the Near East, and South and Central America, saccharification of the starch in cereals - such as barley, corn, or wheat - has long been done by malting the grain. This procedure is followed by the production of alcoholic beverages and food through the action of *Saccharomyces* spp. In the Orient, *Aspergillus* spp. and *Rhizopus* spp. remain in use to make alcoholic beverages and foods, and the same two fungal species are also employed to hydrolyze the proteins of fish, meat, beans, pulses, and some cereals.

Other Fungally Fermented Foods

Some cheeses are made flavorful - following the formation of the curd and its processing - through the action of enzymes of the fungi *Penicillium camem-*

bert (Camembert and Brie) and *Penicillium roqueforti* (Bleu, Gorgonzola, Roquefort, and Stilton). Country-cured hams are produced through fermentation by *Aspergillus* and *Penicillium* spp.; tuna is fermented by *Aspergillus glaucus*, cocoa by *Candida krusei* and *Geotrichum* spp., and peanut presscake by *Neurospora sitophila* (Jay 1986).

Fungal Secondary Metabolites

Fungi produce a large variety of secondary metabolites, but often only when the fungal cells cease active growth. Some of these secondary metabolites are beneficial to humans, whereas others are toxic, and still others may have useful medical effects.

Fungi supply organic acids for industrial uses: citric acid for the food, beverage, pharmaceutical, cosmetic, and detergent industries; itaconic acid for the plastic, paint, and printer's-ink industries; fumaric acid for the paper, resin, fruit juice, and dessert industries (Bigelis and Arora 1992; Zidwick 1992); gluconic acid for the food, beverage, cleaning, and metal-finishing industries; and malic and lactic acids for the food and beverage industries. In addition, several fungi produce rennets for the dairy industry; among these are *Byssoschlamys fulva*, *Candida lipolytica*, *Chlamydomucor oryzae*, *Flammulina velutipes*, *Rhizopus* spp., and *Trametes ostreiformis* (Sternberg 1978).

Certain fungi (especially *Streptomyces* spp.) have proven to be useful as sources of a host of antibiotics that act as inhibitors of bacterial cell-wall synthesis (Oiwa 1992), as antifungal agents (Tanaka 1992), as antiviral agents (Takeshima 1992), and as antiprotozoal and anthelmintic agents (Otoguro and Tanaka 1992). Some also produce antitumor compounds (Komiya and Funayama 1992), cell-differentiation inducers (Yamada 1992), enzyme inhibitors (Tanaka et al. 1992), immunomodulation agents (Yamada 1992), and vasoactive substances (Nakagawa 1992). In addition, fungi have been used to produce herbicides (Okuda 1992), fungicides, and bactericides of plants (Okuda and Tanaka 1992).

A number of secondary metabolites of fungi, however, are toxic to humans and their domestic animals. Aflatoxins are hepatotoxic and carcinogenic; deoxynivalenol is emetic; ergot alkaloids are vasoconstrictive, gangrenous, hemorrhagic, and neurotoxic; zearalenone causes vulvovaginitis in swine; trichothecenes produce vomiting, oral necrosis, and hemorrhage; ochratoxin causes nephrotoxicity; and macrocyclic trichothecenes cause mucosal necrosis (Marasas and Nelson 1987).

Those species of the fungal genus *Claviceps* that grow on cereals (as for example, *Claviceps purpurea*) produce a variety of pharmacologically active compounds with positive and negative effects on humans. Among these are the alkaloids lysergic acid diethylamide (LSD), ergometrine, ergotrienine, ergotamine, ergosinine, ergocristine, ergocornine, ergocristinine,

ergocryptine, and ergocryptinine. Some of these alkaloids are responsible for the disease ergotism, but others are used beneficially – in childbirth, or to treat migraines (Johannsson 1962).

Still other fungi associated with cereals and legumes produce a wide variety of toxins. These have been implicated in aflatoxin and liver cancer in Africa, in esophageal cancer in Africa and Asia, and in endemic nephritis in the Balkans (Stoloff 1987).

A number of fungi (i.e., *Fusarium* and *Gibberella* spp.) produce zearalanol, which exhibits estrogen activity. These estrogen-like compounds are frequent contaminants in cereals and may be responsible for carcinogenesis and precocious sexual development if present in quantity (Schoental 1985).

Aspergillus flavus, which grows on peanuts, soybeans, cereals, and other plants, may produce the hepatocarcinogen aflatoxin and can cause Reye's syndrome in children. *Fusarium* spp., also growing on cereals, can produce trichothecene toxins that cause toxic aleukia (ATA) and *akakabi-byo* ("red mold disease") in Japan.

The commonly cultivated mushroom, *Agaricus bisporus*, may contain phenylhydrazine derivatives that have been found to be weakly mutagenic. Many other edible fungi have shown mutagenic activity (Chauhan et al. 1985); among them is the false morel, *Gyromitra esculenta*, which has been found to contain 11 hydrazines, including gyromitrin – and 3 of these hydrazines are known mutagens and carcinogens (Toth, Nagel, and Ross 1982; Ames 1983; Meier-Bratschi et al. 1983).

In addition, a number of wild fungi contain poisonous molecules that can cause serious illness or death. The amount of poison varies from species to species and from strain to strain within individual species (Benedict and Brady 1966). Also, humans vary in their tolerance of fungal poisons (Simmons 1971).

Fungal toxins produce a variety of biological effects: Amanitin, phallotoxins, and gyromitrin cause kidney and liver damage; coprine and muscarine affect the autonomic nervous system; ibotenic acid, muscimol, psilocybin, and psilocin affect the central nervous system and cause gastrointestinal irritation; indeed, many of these substances and other unknown compounds found in fungi are gastrointestinal irritants (Diaz 1979; Fuller and McClintock 1986).

Several edible fungi, such as *Coprinus atramentarius*, *Coprinus quadrifidus*, *Coprinus variegatus*, *Coprinus insignis*, *Boletus luridus*, *Clitocybe clavipes*, and *Verpa bohemica*, may contain coprine (Hatfield and Schaumburg 1978). Indeed, European *C. atramentarius* may have as much as 160 mg of coprine per kg of fresh fungi. In the human body, coprine is hydrolyzed to l-aminocyclopropanol hydrochloride (ACP), which acts like disulfuram, a synthetic compound known as antabuse and used to treat chronic alcoholics. Antabuse and ACP irre-

versibly inhibit acetaldehyde dehydrogenase and prevent the catabolism of ethanol. Thus, coprine plus ethanol leads to severe intoxication when alcoholic beverages are drunk after eating coprine-containing fungi (Hatfield and Schaumburg 1978; Hatfield 1979).

In addition, many mushrooms contain the enzyme thiaminase, which may destroy the vitamin thiamine, leading to thiamine deficiency (Wakita 1976) – especially when the mushrooms are eaten in quantity (Rattanapanone 1979). Several *Russula* spp. may contain indophenolase, which can also be harmful to humans if eaten in large amounts (Romagnesi 1967).

Humans can become allergic to fungi (Koivikko and Savolainen 1988). Moreover, eating fava beans with mushrooms that are rich in tyrosinase may enhance the medical effect of the fava beans – known as favism – because the tyrosinase catalyzes the conversion of L-DOPA to L-DOPA-quinone (Katz and Schall 1986).

Magico-Religious Use of Fungi

As early as the eighteenth century, according to travelers' reports, *Amanita muscaria*, known as the "fly agaric," was eaten by several tribal groups (Chukchi, Koryak, Kamchadal, Ostyak, and Vogul) in eastern Siberia as an intoxicant and for religious purposes. Species of *Panaeolus*, *Psilocybe*, and *Stropharia* also contain hallucinogens. These fungi were eaten by the Aztecs and the Maya – and are still consumed by *curanderos* in some Mexican tribes – to produce hallucinations for religious purposes, to derive information for medical treatment, and to locate lost objects (Diaz 1979).

Sheldon Aaronson

This work was funded, in part, by research awards from PSC/CUNY and the Ford Foundation.

Bibliography

- Aalbersberg, W. G. L., C. E. A. Lovelace Madhaji, and S. V. Parekenson. 1988. Davuke, the traditional Fijian method of pit preservation of staple carbohydrate foods. *Ecology of Food and Nutrition* 21: 173–80.
- Abercrombie, J. 1779. *The garden mushroom, its nature and cultivation*. London.
- Adams, C. F. 1988. *Nutritive value of American foods*. Agriculture Handbook No. 456. Washington, D.C.
- Ahmed, A. A., M. A. Mohamed, and M. A. Hani. 1981. Libyan truffles: Chemical composition and toxicity. *Mushroom Science* 11: 833–42.
- Ainsworth, G. C. 1976. *Introduction to the history of mycology*. Cambridge and New York.
- Alcorn, J. B. 1984. *Huastec Mayan ethnobotany*. Austin, Tex.
- Alsheikh, A., and J. M. Trappe. 1983. Desert truffles: The

- genus *Tirmania*. *Transactions of the British Mycology Society* 81: 83-90.
- Ames, B. N. 1983. Dietary carcinogens and anticarcinogens. *Science* 221: 1256-64.
- Ames, B. N., J. McCann, and F. Yamaski. 1975. Methods for detecting carcinogens and mutagens with the *Salmonella*/mammalian microsome mutagenicity test. *Mutation Research* 31: 347-64.
- André, J. 1985. *Les noms de plantes dans la Rome Antique*. Paris.
- Arenas, P. 1981. *Ethnobotaneca Lengua-maskoy*. Buenos Aires.
- Arnason, T., R. J. Hebda, and T. Johns. 1981. Use of plants for food and medicine by native peoples of eastern Canada. *Canadian Journal of Botany* 59: 2189-325.
- Atal, C. K., B. K. Bhat, and T. N. Kaul. 1978. *Indian mushroom science I*. Globe, Ariz.
- Bano, Z. 1978. The nutritive value of mushrooms. In *Indian mushroom science I*, 473-87. Globe, Ariz.
- Bano, Z., and S. Rajarathnam. 1982. *Pleurotus* mushrooms: A nutritious food. In *Tropical mushrooms*, ed. S. T. Chang and T. H. Quimio, 363-80. Hong Kong.
- Barrau, J. 1962. Ph.D. thesis, presented to the Faculté des Sciences de Marseille.
- Bels, P. J., and S. Pataragetvit. 1982. Edible mushrooms in Thailand, cultivated by termites. In *Tropical mushrooms*, ed. S. T. Chang and T. H. Quimio, 445-61. Hong Kong.
- Benedict, R. G., and L. R. Brady. 1966. Occurrence of *Amanita* toxins in American collections of deadly *Amanitas*. *Journal of Pharmaceutical Science* 55: 590-3.
- Beuchat, L. R. 1983. Indigenous fermented foods. In *Biotechnology*, ed. H. J. Rehm, and G. Reed, 8 vols., 5: 477-528. Weinheim, Germany.
- Bigelis, R. 1992. Food enzymes. In *Biotechnology of filamentous fungi*, ed. D. B. Finkelstein and C. Ball, 361-415. Boston.
- Bigelis, R., and D. K. Arora. 1992. Organic acids of fungi. In *Handbook of applied mycology*, Vol. 4, ed. D. K. Arora, R. P. Elander, and K. G. Mukerji, 357-76. New York.
- Blyth, R. H. 1973. Mushrooms in Japanese verse. *Transactions of the Asiatic Society, Japan* 11: 1-14.
- Bo, L., and B. Yun-sun. 1980. *Fungi Pharmacopoeia (Sinica)*. Oakland, Calif.
- Bose, S. R., and A. B. Bose. 1940. An account of edible mushrooms of India. *Science and Culture* 6: 141-9.
- Bramley, P. M., and A. Mackenzie. 1992. Carotenoid biosynthesis and its regulation in fungi. *Handbook of applied mycology*, Vol. 4, ed. D. K. Arora, R. P. Elander, and K. G. Mukerji, 401-44. New York.
- Buller, A. H. R. 1914-16. The fungus lore of the Greeks and Romans. *Transactions of the British Mycological Society* 5: 21-66.
- Burkill, T. H. 1935. *A dictionary of the economic products of the Malay Peninsula*. London.
- Cao, J., et al. 1991. A new wild edible fungus - *Wynnella silvicola*. *Edible Fungi of China* 10: 27.
- Casalichchio, G., C. Paoletti, A. Bernicchia, and G. Gooi. 1975. Ricerche sulla composizione aminoacidica di alcuni funghi. *Micologica Italiana* 1: 21-32.
- Chang, S. T., and W. A. Hayes. 1978. *The biology and cultivation of edible mushrooms*. London.
- Chang, S. T., and P. G. Miles. 1987. Historical record of the early cultivation of *Lentinus* in China. *Mushroom Journal in the Tropics* 7: 31-7.
- Chao Ken, N. 1980. The knowledge and usage of fungus in ancient China. *Acta Microbiologica Sinica* 7: 174-5.
- Chauhan, Y., D. Nagel, M. Gross, et al. 1985. Isolation of N2-(gamma-[+glutamyl]-4carboxyphenyl hydrazine) in the cultivated mushroom *Agaricus bisporus*. *Journal of Agricultural Food Chemistry* 33: 817-20.
- Chavan, J. K., and S. S. Kadam. 1989. Nutritional improvement of cereal by fermentation. *Critical Reviews in Food Science and Nutrition* 28: 349-400.
- Chen, G. 1989a. Wild edible and medical fungi resources in Shenyang City. *Edible Fungi of China* 36: 30.
- 1989b. Wild edible and medical fungi resources in Shenyang City. *Edible Fungi of China* 37: 25-6.
- Chestnut, V. K. 1902. Plants used by the Indians of Mendocino County, California. *U.S. National Herbarium*: 294-422.
- Colenso, W. 1881. On the vegetable food of the ancient New Zealanders before Cook's visit. *Transactions and Proceedings of the New Zealand Institute* 13: 3-19.
- Cravioto, O. R., Y. O. Cravioto, H. G. Massieu, and G. T. Guzman. 1955. El pozol, forma indígena de consumir el maíz en el sureste de Mexico y su aporte de nutrientes a la dieta. *Ciencia Mexicana* 15: 27-30.
- Cribb, A. B., and J. W. Cribb. 1975. *Wild foods in Australia*. Sydney.
- Crisan, E. V., and A. Sands. 1978. Nutritional value. In *The biology and cultivation of edible mushrooms*, ed. S. T. Chang and W. A. Hayes, 137-81. New York.
- Davis, E. W., and J. A. Yost. 1983. The ethnobotany of the Waorani of eastern Ecuador. *Botanical Museum Leaflets Harvard University* 3: 159-217.
- Deonna, W., and M. Renard. 1961. *Croyances et superstitions de table dans la Romantique*. Brussels.
- Diaz, J. L. 1979. Ethnopharmacology and taxonomy of Mexican psychodelic plants. *Journal of Psychedelic Drugs* 11: 71-101.
- Dickson, V. 1971. *Forty years in Kuwait*. London.
- Dirar, H. A. 1993. *The indigenous fermented foods of the Sudan: A study in African food and nutrition*. Wallingford, England.
- FAO (Food and Agriculture Organization of the United Nations). 1970. *Amino acid content of foods*. FAO Nutritional Studies No. 24. Rome.
- Ferrante, M. P., and A. Fiechter. 1983. *Production and feeding of single cell protein*. London.
- Findlay, W. P. K. 1982. *Fungi, folklore, fiction, and fact*. Eureka, Calif.
- Forbes, R. J. 1967. The beginnings of technology and man. *Technology in Western civilization*, ed. M. Kranzberg and C. W. Pursell, Jr., 11-47. New York.
- Fuller, T. C., and E. McClintock. 1986. *Poisonous plants of California*. Berkeley, Calif.
- Gade, D. W. 1975. *Plants, man and the land in the Vulcanota Valley of Peru*. The Hague.
- Gallois, A., B. Gross, D. Langlois, et al. 1990. Flavor compounds produced by lignolytic basidiomycetes. *Mycological Research* 94: 494-504.
- Gandhi, S. R., and J. D. Weete. 1991. Production of the polyunsaturated fatty acids arachidonic acid and eicosentaenoic acid by the fungus *Pythium ultimum*. *Journal of General Microbiology* 137: 1825-30.
- Gardner, W. A., and C. W. McCoy. 1992. Insecticides and herbicides. In *Biotechnology of filamentous fungi*, ed. D. B. Finkelstein and C. Ball, 335-59. Boston, Mass.
- Geller, J. R. 1989. Recent excavations at Hierakonpolis and their relevance to Predynastic pyrotechnology and settlement. *Cahiers de recherches de l'Institut de papyrologie et d'egyptologie de Lille* 11: 41-52.

1992. From prehistory to history: Beer in Egypt. In *The followers of Horus*, ed. R. Friedman and B. Adams. 19–26. Oxford.
- Goldberg, L., and J. M. Thorp. 1946. A survey of vitamins in African foodstuffs. VI. Thiamin, riboflavin and nicotinic acid in sprouted and fermented cereal foods. *South African Journal of Medical Science* 11: 177–85.
- Goody, Jack. 1982. *Cooking, cuisine, and class: A study in comparative sociology*. Cambridge and New York.
- Grieve, M. 1925. *Fungi as food and in medicine*. Tamworth, England.
- Griffin, D. H. 1981. *Fungal physiology*. New York.
- Gruss, J. 1928. *Saccharomyces winlocki*, die Hefe aus den Pharaonengräbern. *Allgemeine Brauer und Hopfen Zeitung* 237: 1340–1.
- Guerra, F. 1967. Mexican Phantastica – a study of the early ethnobotanical sources on hallucinogenic drugs. *British Journal of Addiction* 62: 171–87.
- Guo, W. 1992. Resources of wild edible fungi in Tibet, China. *Edible Fungi of China* 11: 33–4.
- Hamil, F. B., and M. U. Chiltonkey. 1975. *Cherokee plants*. Sylva, N.C.
- Hart, J. A. 1979. The ethnobotany of the Flathead Indians of western Montana. *Botanical Museum Leaflet* 27: 261–307.
- Hatfield, G. M. 1979. Toxic mushrooms. In *Toxic plants*, ed. A. Douglas Kinghorn, 7–58. New York.
- Hatfield, G. M., and J. P. Schaumberg. 1978. The disulfuram-like effects of *Coprinus atramentarius* and related mushrooms. In *Mushroom poisoning, diagnosis and treatment*, ed. B. H. Rumack and E. Salzman, 181–6. West Palm Beach, Fla.
- Hawksworth, D. L., B. C. Sutton, and G. A. Ainsworth. 1983. *Ainsworth & Bisby's dictionary of the fungi*. Kew, England.
- Hayes, W. A. 1987. Edible mushrooms. In *Food and beverage mycology*, ed. L. R. Beuchat, 355–90. New York.
- Haytowitz, D. B., and R. H. Matthews. 1984. *Composition of foods*. Washington, D.C.
- Herrera, F. L. 1934. Botanica ethnolic. *Revista del Museo Nacional de Lima Peru* 3: 37–62.
- Herrero, P. 1984. *La thérapeutique mesopotamienne*. Mémoire No. 48. Paris.
- Hesseltine, C. W. 1985. Fungi, people, and soybeans. *Mycologia* 77: 505–25.
- Hiepkö, P., and W. Schultze-Motel. 1981. *Floristische und ethnobotanische Untersuchungen im Eipomektal. Irian Jaya*. Berlin.
- Hohl, H. R. 1987. Cytology and morphogenesis of fungal cells. *Progress in Botany* 49: 13–28.
- Holland, H. L. 1992. *Bioconversions*, ed. D. B. Finkelstein and C. Ball, 157–87. Boston.
- Holland, H. L., F. M. Brown, J. A. Rao, and P. C. Chenchiah. 1988. Synthetic approaches to the prostaglandins using microbial biotransformation. *Developments in Industrial Microbiology* 29: 191–5.
- Holter, U. 1988. Food consumption of camel nomads in the north-west Sudan. *Ecology of Food and Nutrition* 21: 95–115.
- Hu, L., and L. Zeng. 1992. Investigation on wild edible mushroom resources in Wanxian County, Sichuan Province. *Edible Fungi of China* 11: 35–7.
- Humfeld, H. 1948. The production of mushroom mycelium (*Agaricus campestris*) in submerged culture. *Science* 107: 373.
- Imai, S. 1938. Studies on the Agaricaceae of Hokkaido. *Journal of the Faculty of Agriculture, Hokkaido Imperial University* 43: 359–63.
- Irvine, F. R. 1952. Supplementary and emergency food plants of West Africa. *Economic Botany* 6: 23–40.
- Irving, F. R. 1957. Wild and emergency foods of Australian and Tasmanian aborigines. *Oceania* 28: 113–42.
- Ito, T. 1917. *Collybia nameko* sp. nov. a new edible fungus of Japan. *Japan Academy Proceedings* 5: 145–7.
1978. Cultivation of *Lentinus edodes*. In *The biology and cultivation of edible mushrooms*, ed. S. T. Chang and W. A. Hayes, 461–73. New York.
- Jay, J. M. 1986. *Modern food microbiology*. New York.
- Johannsson, M. 1962. Studies in alkaloid production by *Claviceps purpurea*. *Symbolae Botanicae Upsaliensis* 17: 1–47.
- Johnson, C. R. 1862. *The useful plants of Great Britain*. London.
- Katz, S. H., and J. J. Schall. 1986. Favism and malaria: A model of nutrition and biocultural evolution. In *Plants in indigenous medicine and diet*, ed. N. L. Elkin, 211–28. Bedford Hills, N.Y.
- Kaul, T. N. 1981. Common edible mushrooms of Jammu and Kashmir. *Mushroom Science* 11: 79–82.
- Kaul, T. N., and J. L. Kachroo. 1974. Common edible mushrooms of Jammu and Kashmir. *Journal of the Bombay Natural History Society* 71: 26–31.
- Kempler, G. M. 1983. Production of flavor compounds by microorganisms. *Advances in Applied Microbiology* 29: 29–51.
- Kerwin, J. L., and N. D. Duddles. 1989. Reassessment of the roles of phospholipids in sexual reproduction by sterol-auxotrophic fungi. *Journal of Bacteriology* 171: 3829–31.
- Klaui, H. 1982. Industrial and commercial use of carotenoids. *Carotenoid biochemistry*, ed. G. Britton and T. W. Goodwin, 309–28. Oxford.
- Ko, H. 1966. *Alchemy, medicine, religion in the China of A.D. 320: The Nei Pieu of Ko Hung (Polo-pu-tzu)*, Trans. James. R. Ware. Cambridge, Mass.
- Kodama, K., and K. Yoshizawa. 1977. Sake. In *Alcoholic beverages*, ed. A. H. Rose, 423–75. London.
- Koivikko, A., and J. Savolainen. 1988. Mushroom allergy. *Allergy* 43: 1–10.
- Komiyama, K., and S. Funayama. 1992. Antitumor agents. In *The search for bioactive compounds from microorganisms*, ed. S. Omura, 79–103. New York.
- Konno, K., K. Hayano, H. Shirakama, et al. 1982. Clitidine, a new toxic pyridine nucleoside from *Clitocybe acromelalgae*. *Tetrahedron* 38: 3281–4.
- Konno, K., H. Shirahama, and T. Matsumoto. 1984. Clithioneine, an amino acid betaine from *Clitocybe acromelalgae*. *Phytochemistry* 23: 1003–6.
- Kuhnlein, H. V., and N. J. Turner. 1991. *Traditional plant foods of Canadian indigenous peoples*. Philadelphia, Pa.
- Kutsal, T., and M. T. Ozbas. 1989. Microbial production of vitamin B₂ (riboflavin). In *Biotechnology of vitamins, pigments and growth factors*, ed. E. T. Vandamme, 149–66. London.
- Lee, R. B. 1979. *The 'Kung San*. Cambridge and New York.
- Levai, J. 1984. Nutritional and utilizable value of some cultivated mushrooms. *Mushroom Science* 12: 295–304.
- Lewis, D. H., and D. C. Smith. 1967. Sugar alcohols (polyols) in fungi and green plants. I. Distribution, physiology and metabolism. *New Phytologist* 66: 143–84.
- Li, G. S. F., and S. T. Chang. 1982. Nutritive value of *Volvariella volvacea*. In *Tropical mushrooms*, ed. S. T. Chang and T. H. Quimio, 199–219. Hong Kong.
- Lin, C.-F., and H. Iizuka. 1982. Production of extracellular pigment by a mutant of *Monascus koaling* sp. nov. *Applied Environmental Microbiology* 43: 671–6.

- Lincoff, G. H. 1984. *Field guide to North American mushrooms*. New York.
- Lipp, F. J. 1991. *The Mixe of Oaxaca, religion, ritual and healing*. Austin, Tex.
- Litchfield, J. H. 1967. Submerged culture of mushroom mycelium. In *Microbial technology*, ed. H. J. Peppler, 107-44. New York.
- Litchfield, J. H., V. G. Vely, and R. C. Overbeck. 1963. Nutrient content of morel mushroom mycelium: Amino acid composition of the protein. *Journal of Food Science* 28: 741-3.
- Liu, B. 1958. The primary investigation on utilization of the fungi by ancient Chinese. *Shansi Normal College Journal* 1: 49-67.
1974. *The gasteromycetes of China*. Vaduz, Liechtenstein.
- Liu, P. 1993. Introduction of a valuable edible fungus from Yunnan - *Lyophyllum shimeii* (Kawam.) Hongo. *Edible Fungi of China* 12: 29.
- Liu, S. 1991. Discussion on *Ganoderma lucidum* in ancient traditional Chinese medical books. *Edible Fungi of China* 10: 37-8.
- Lou, L. H. 1982. Cultivation of *Auricularia* on logs in China. In *Tropical mushrooms*, ed. S. T. Chang and T. H. Quimio, 437-41. Hong Kong.
- Lowe, D. A. 1992. Fungal enzymes. In *Handbook of applied mycology*, Vol. 4., ed. D. K. Arora, R. P. Elander, and K. G. Mukerji, 681-706. New York.
- Maciarello, M. J., and A. O. Tucker. 1994. Truffles and truffle volatiles. In *Spices, herbs and edible fungi*, ed. G. Charalambous, 729-39. Amsterdam.
- Majno, G. 1975. *The healing hand: Man and wound in the ancient world*. Cambridge, Mass.
- Mao, X. 1991. A trip to Yunnan - edible mushroom kingdom. *Edible Fungi of China* 10: 33-4.
- Mapes, C., G. Guzman, and J. Cabellero. 1981. *Etnomicologia Purepbecha*. Lima.
- Marasas, W. F. O., and P. E. Nelson. 1987. *Mycotoxigenology*. University Park, Pa.
- Martinez, A. M. A., E. Perez-Silva, and E. Agurre-Accosta. 1983. Etnomicologia y exploraciones micologicas en la sierra norte de Puebla. *Boletin de la Sociedad Medicina Microbiologica* 18: 51-63.
- Martini, A. E. V., M. W. Miller, and A. Martini. 1979. Amino acid composition of whole cells of different yeasts. *Journal of Agriculture and Food Chemistry* 27: 982-4.
- Mau, J., R. B. Beelman, and G. R. Ziegler. 1994. Aroma and flavor components of cultivated mushrooms. In *Spices, herbs and edible fungi*, ed. G. Charalambous, 657-83. Amsterdam.
- McConnell, J. E. W., and W. B. Esselen. 1947. Carbohydrates in cultivated mushrooms. *Food Research* 12: 118-21.
- Mead, G. K. 1972. *Ethnobotany of the California Indians*. Ethnology Series No. 30. Greeley, Colo.
- Meier-Bratschi, A., B. M. Carden, J. Luthy, et al. 1983. Methylation of deoxyribonucleic acid in the rat by the mushroom poison, gyromitrin. *Journal of Agriculture and Food Chemistry* 31: 1117-20.
- Michel, R. H., P. E. McGovern, and V. R. Badler. 1992. Chemical evidence for ancient beer. *Nature* 360: 24.
- Miller, S. A. 1968. Nutritional factors in single-cell protein. In *Single-cell protein*, ed. R. I. Mateles and S. R. Tannenbaum, 79-89. Cambridge, Mass.
- Molitoris, H. P., J. L. Van Etten, and D. Gottlieb. 1968. Alterbedingte Änderungen der Zusammensetzung und des Stoffwechsels bei Pilzen. *Mushroom Science* 7: 59-67.
- Montes, M. D. 1961. Sobre el uso de hierbas en la medicina popular de Santander (Columbia). *Thesaurus* 16: 719-29.
- Monthoux, O., and K. Lündstrom-Baudais. 1979. Polyporacées des sites néolithiques de Clairvaux et Charavines (France). *Candollea* 34: 153-66.
- Morris, B. 1987. *Common mushrooms of Malawi*. Oslo.
- Murata, K. 1985. Formation of antioxidants and nutrients in tempe. In *Non-salted soybean fermentation*. Asian Symposium, International Committee on Economic and Applied Microbiology. Tsukuba, Japan.
- Nakagawa, A. 1992. Vasoactive substances. In *The search for bioactive compounds from microorganisms*, ed. S. Omura, 198-212. New York.
- Nakajima, N., K. Ichikawa, M. Kamada, and E. Fujita. 1961. Food chemical studies on 5'-ribonucleotides. Part I. On the 5'-ribonucleotides in various stocks by ion exchange chromatography. *Journal of the Agricultural Chemistry Society, Japan* 9: 797-803.
- Needham, J., and L. Gwei-Djin. 1983. Physiological alchemy. In *Science and Civilization in China. Chemistry and Chemical Technology, Spagyric Discovery and Invention*. Part V: 140. Cambridge.
- Nes, W. R. 1977. The biochemistry of plant sterols. *Advances in Lipid Research* 15: 233-324.
- Nishitoba, T., H. Sato, and S. Sakamura. 1988. Bitterness and structure relationship of the triterpenoids from *Ganoderma lucidum* (Reishi). *Agricultural Biological Chemistry* 52: 1791-5.
- Oakley, K. 1962. On man's use of fire, with comments on tool-making and hunting. In *The social life of early man*, ed. S. Washburn, 176-93. Chicago.
- O'Donnell, K., and S. W. Peterson. 1992. Isolation, preservation, and taxonomy. In *Biotechnology of filamentous fungi*, ed. D. B. Finkelstein and C. Ball, 7-33. Boston.
- Oiwa, R. 1992. Antibacterial agents. *The search for bioactive compounds from microorganisms*, ed. S. Omura, 1-29. New York.
- Okuda, S. 1992. Herbicides. In *The search for bioactive compounds from microorganisms*, ed. S. Omura, 224-36. New York.
- Okuda, S., and Y. Tanaka. 1992. Fungicides and antibacterial agents. *The search for bioactive compounds from microorganisms*, ed. S. Omura, 213-23. New York.
- Otoguro, K., and H. Tanaka. 1992. Antiparasitic agents. In *The search for bioactive compounds from microorganisms*, ed. S. Omura, 63-78. New York.
- Pacioni, G., C. Bellina-Agostinone, and M. D'Antonio. 1990. Odour composition of the *Tuber melanoporum* complex. *Mycological Research* 94: 201-4.
- Pan, X. 1993a. Medicinal edible fungi resources in the forest region in Heilongjiang Province. *Edible Fungi of China* 12: 38-9.
- 1993b. Medicinal edible fungi resources in Heilongjiang Province. *Edible Fungi of China* 12: 25-7.
- Paredes-Lopez, C., and G. I. Harry. 1988. Food biotechnology review: Traditional solid-state fermentations of plant raw materials application, nutritional significance, and future prospects. *Reviews in Food Science and Nutrition* 27: 159-87.
- Pegler, D. N., and G. D. Pearce. 1980. The edible mushrooms of Zambia. *Kew Bulletin* 35: 475-92.
- Peters, R., and E. M. O'Brien. 1984. On hominid diet before fire. *Current Anthropology* 25: 358-60.
- Petrie, W. M., and J. E. Quibell. 1896. *Nagada and Ballas*. London.
- Pearce, G. D. 1981. Zambian mushrooms - customs and folklore. *British Mycological Society Bulletin* 139-42.
- Pöder, R., U. Peintner, and T. Pümpel. 1992. Mykologische Untersuchungen an den Pilz-Beifunden der Gletscher-

- mumie vom Hauslabjoch. *Der Mann im Eis*, ed. F. Hopfel, W. Platzler, and K. Spindler, 313–20. Innsbruck.
- Prance, G. T. 1972. An ethnobotanical comparison of four tribes of Amazonian Indians. *Amazonica* 1: 7–27.
1983. The ethnobotany of Amazon Indians: A rapidly disappearing source of botanical knowledge for human welfare. *Bulletin Botanical Survey of India* 25: 148–59.
- Purkayastha, R. P. 1978. Indian edible mushrooms – a review. In *Indian mushroom science I*, ed. C. K. Atal, B. K. Bhat, and T. N. Kaul, 351–71. Srinagar, India.
- Purkayastha, R. P., and A. Chandra. 1976. Amino acid composition of the protein of some edible mushrooms grown in synthetic medium. *Journal of Food Science and Technology* 13: 86–9.
- Radwan, S. S., and A. H. Soliman. 1988. Arachidonic acid from fungi utilizing fatty acids with shorter chains as sole sources of carbon and energy. *Journal of General Microbiology* 134: 387–93.
- Romain, P. 1981. Essai de mycogastronomie. *Revue de Mycologie, Supplement* 12: 4–18, 29–38, 75–82.
- Rattanapanone, V. 1979. Antithiamin factor in fruit, mushroom and spices. *Chiang Mai Medical Bulletin* (January): 9–16.
- Reagan, A. B. 1929. Plants used by the White Mountain Apache Indians of Arizona. *Wisconsin Archaeologist* 8: 143–61.
- Reddy, N. R., M. D. Person, and D. K. Salunkhe. 1986. Idli. In *Legume-based fermented foods*, ed. N. R. Reddy, M. D. Pierson, and D. K. Salunkhe, 145–60. Boca Raton, Fla.
- Reinking, O. A. 1921. Philippine edible fungi. In *Minor products of Philippine forests*, ed. W. H. Brown, 103–47. Manila.
- Robinson, R. F., and R. S. Davidson. 1959. The large scale growth of higher fungi. *Advances in Applied Microbiology* 1: 261–78.
- Rolfe, R. T., and F. W. Rolfe. 1925. *The romance of the fungus world*. London.
- Romagnesi, H. 1967. *Les Russules d'Europe et d'Afrique du Nord*. Paris.
- Sabban, F. 1986. Court cuisine in fourteenth-century imperial China: Some culinary aspects of Hu Sihui's Yingshan Zengyao. *Food and Foodways* 1: 161–96.
- Saffirio, L. 1972. Food and dietary habits in ancient Egypt. *Journal of Human Evolution* 1: 197–305.
- Saggs, H. W. F. 1962. *The greatness that was Babylon*. New York.
- Said, H. M., R. E. Elahie, and S. K. Hamarneh. 1973. *Al-Biruni's book on pharmacy and materia medica*. Karachi.
- Sakaguchi, K. 1972. Development of industrial microbiology in Japan. *Proceedings of the International Symposium on Conversion and Manufacture of Foodstuffs by Microorganisms*, 7–10. Tokyo.
- Schnell, R. 1957. *Plantes alimentaires et vie agricole de l'Afrique Noire*. Paris.
- Schoental, R. 1985. Zearalenone, its biological and pathological, including carcinogenic effects in rodents: Implications for humans. *Fifth meeting on mycotoxins in animal and human health*, ed. M. O. Moss and M. Frank, 52–72. Edinburgh.
- Schreier, P. 1992. Bioflavours: An overview. In *Bioformation of flavorus*, ed. R. L. S. Patterson, B. V. Charlwood, G. MacLeod, and A. A. Williams, 1–20. Cambridge.
- Schultes, R. E. 1937. Teonanacatl: The narcotic mushroom of the Aztecs. *American Anthropologist* 42: 424–43.
- Schultes, R. E., and A. Hoffmann. 1979. *Plants of the gods*. New York.
- Scudder, T. 1971. *Gathering among African woodland savannab cultivators. A case study: The Gwembe Tonga*. Manchester, England.
- Semerdzieva, M., M. Wurst, T. Kozu, and J. Gartz. 1986. Psilocybin in Fruchtkörpern von *Inocybe aeruginascens*. *Planta Medica* 2: 83–5.
- Sensarma, P. 1989. *Plants in the Indian Puranas*. Calcutta.
- Sharples, R. W., and D. W. Minter. 1983. Theophrastus on fungi: Inaccurate citations in Athanaeus. *Journal of Hellenic Studies* 103: 154–6.
- Shaw, D. E. 1984. *Microorganisms in Papua New Guinea*. Research Bulletin No. 33. Port Moresby.
- Shaw, R. 1966. The polyunsaturated fatty acids of microorganisms. *Advances in Lipid Research* 4: 107–74.
- Shinmen, Y., S. Shimazu, K. Akimoto, et al. 1989. Production of arachidonic acid by *Mortierella* fungi. *Applied Microbiology and Biotechnology* 31: 11–16.
- Simmons, D. M. 1971. The mushroom toxins. *Delaware Medical Journal* 43: 177–87.
- Singer, R. 1961. *Mushrooms and truffles*. London.
1978. Hallucinogenic mushrooms. In *Mushroom poisoning: Diagnosis and treatment*, ed. B. H. Rumack and E. Salzman, 201–14. West Palm Beach, Fla.
- Singh, T. B., and K. C. Chuneekar. 1972. *Glossary of vegetable drugs in Brhatrayi*. Varanasi, India.
- Sinskey, A. J., and C. A. Batt. 1987. Fungi as a source of protein. In *Food and beverage mycology*, ed. L. R. Beuchat, 435–71. New York.
- Stanton, W. R. 1985. Food fermentation in the tropics. In *Microbiology of fermented foods*, Vol. 2, ed. B. J. B. Wood, 193–211. London.
- Steinkraus, K. H. 1983. Traditional food fermentation as industrial resources. *Acta Biotechnologia* 3: 3–12.
1985. Production of vitamin B-12 in tempe. In *Non-salted soybean fermentation*. Asian Symposium, International Committee on Economic and Applied Microbiology, 68. Tsukuba, Japan.
- Sternberg, M. 1978. Microbial rennets. *Advances in Applied Microbiology* 24: 135–57.
- Stijve, T., J. Klan, and T. W. Kuyper. 1985. Occurrence of psilocybin and baeocystin in the genus *Inocybe* (Fr.) Fr. *Persoonia* 2: 469–73.
- Stoloff, L. 1987. Carcinogenicity of aflatoxin. *Science* 237: 1283.
- Strong, F. M. 1974. Toxicants occurring naturally in foods. *Nutrition Reviews* 32: 225–31.
- Stuart, D. E. 1977. Seasonal phases in Ona subsistence, territorial distribution and organization. Implications for the archeological record. In *For theory building in archaeology*, ed. L. R. Binford, 251–83. New York.
- Takeshima, H. 1992. Antiviral agents. In *The search for bioactive compounds from microorganisms*, ed. S. Omura, 45–62. New York.
- Tanaka, H., K. Kawakita, N. Imamura, et al. 1992. General screening of enzyme inhibitors. In *The search for bioactive compounds from microorganisms*, ed. S. Omura, 117–80. New York.
- Tanaka, N. 1890. On hatsudake and akahatsu, two species of edible fungi. *Botanical Magazine* 4: 2–7.
- Tanaka, Y. 1992. Antifungal agents. In *The search for bioactive compounds from microorganisms*, ed. S. Omura, 30–44. New York.
- Tanaka, Y., A. Hasegawa, S. Yamamoto, et al. 1988. World-wide contamination of cereals by the *Fusarium* mycotoxins nivalenol and zearalenone. I. Surveys of 19

- countries. *Journal of Agricultural and Food Chemistry* 36: 979-83.
- Tanaka, Y., and S. Okuda. 1992. Insecticides, acaricides and anticoccidial agents. In *The search for bioactive compounds from microorganisms*, ed. S. Omura, 237-62. New York.
- Tee, E. S. 1992. Carotenoids and retinoids in human nutrition. *Reviews in Food Science and Nutrition* 31: 103-63.
- Terrell, E. E., and L. R. Batra. 1982. *Zizania latifolia* and *Ustilago esculenta*, a grass-fungus association. *Economic Botany* 36: 274-85.
- Thrower, L. B., and Y.-S. Chan. 1980. *Gau sun*: A cultivated host parasite combination from China. *Economic Botany* 34: 20-6.
- Toth, B. 1983. Carcinogens in edible mushrooms. *Carcinogens and Mutagens in the Environment* 3: 99-108.
- Toth, B., D. Nagel, and A. Ross. 1982. Gastric tumorigenesis by a single dose of 4-(hydroxymethyl)benzenediazonium ion of *Agaricus bisporus*. *British Journal of Cancer* 46: 417-22.
- Tounefort, J. de. 1707. Observations sur la naissance et sur la culture des champignons. *Mémoires de l'Académie Royale des Sciences*, 58-66.
- Tung, T. C., P. C. Huang, and H. O. Li. 1961. Composition of foods used in Taiwan. *Journal of the Formosan Medical Association* 60: 973-1005.
- Turner, N. J. 1978. *Food plants of British Columbia Indians. Part II-interior peoples*. Victoria.
- Turner, N. J., H. V. Kuhnlein, and K. N. Egger. 1987. The cottonwood mushroom (*Tricholoma populinum* Lange): A food resource of the Interior Salish Indian peoples of British Columbia. *Canadian Journal of Botany* 65: 921-7.
- Umezawa, C., and T. Kishi. 1989. Vitamin metabolism. In *Metabolism and physiology of yeasts*, ed. A. H. Rose and J. S. Harrison, 457-88. London.
- Usher, G. 1974. *A dictionary of plants used by man*. New York.
- Vandamme, E. J. 1989. Vitamins and related compounds via microorganisms; a biotechnological view. In *Biotechnology of vitamins, pigments and growth factors*, ed. E. G. Vandamme, 1-11. London.
- Verma, R. N., and T. G. Singh. 1981. Investigation on edible fungi in the north eastern hills of India. *Mushroom Science* 11: 89-99.
- Wakita, S. 1976. Thiamine-distribution by mushrooms. *Science Report, Yokohama National University* 2: 39-70.
- Wang, Y. C. 1985. Mycology in China with emphasis on review of the ancient literature. *Acta Mycologica Sinica* 4: 133-40.
- Wassef, M. K. 1977. Fungal lipids. *Advances in Lipid Research* 15: 159-232.
- Wasson, R. G. 1975. Mushrooms and Japanese culture. *Transactions of the Asiatic Society of Japan* (Third Series) 11: 5-25.
- Watling, R., and M. R. D. Seaward. 1976. Some observations on puffballs from British archaeological sites. *Journal of Archaeological Science* 3: 165-72.
- Weete, J. D. 1974. *Fungal lipid chemistry*. New York.
1989. Structure and function of sterols in fungi. *Advances in Lipid Research* 23: 115-67.
- Weete, J. D., M. S. Fuller, M. Q. Huang, and S. Gandhi. 1989. Fatty acids and sterols of selected hyphochytridomycetes. *Experimental Mycology* 13: 183-95.
- Winlock, H. E. 1973. *The tomb of Queen Meryet-Amun at Thebes*. New York.
- Wong, H. C., and P. E. Koehler. 1981. Production and isolation of an antibiotic from *Monascus purpureus* and its relationship to pigment production. *Journal of Food Science* 46: 589-92.
- Yamada, H. 1992. Immunomodulators. In *The search for bioactive compounds from microorganisms*, ed. S. Omura, 171-97. New York.
- Yamada, H., S. Shimizu, and Y. Shinmen. 1987. Production of arachidonic acid by *Mortierella elongata* IS-5. *Agricultural and Biological Chemistry* 51: 785-90.
- Yamada, H., S. Shimizu, Y. Shinmen, et al. 1988. Production of arachidonic acid and eicosapentenoic acid by microorganisms. In *Proceedings of the World Conference on the Biotechnology of Fats Oils Industry*. American Oil Chemists' Society, 173-7. Champaign, Ill.
- Yanovsky, E. 1936. *Food plants of the North American Indians*. United States Department of Agriculture Misc. Pub. No. 237. Washington, D.C.
- Yen, D. E., and H. G. Gutierrez. 1976. The ethnobotany of the Tasaday: I. The useful plants. In *Further studies on the Tasaday*, ed. D. E. Yen and J. Nance, 97-136. Makati, Rizal.
- Yokotsuka, T. 1985. Fermented protein foods in the Orient with emphasis on shoyu and miso in Japan. In *Microbiology of fermented foods*, Vol. 1., ed. B. J. B. Wood, 197-247. London.
1986. Soy sauce biochemistry. *Advances in Food Research* 30: 195-329.
- Yokoyama, K. 1975. Ainu names and uses for fungi, lichens and mosses. *Transactions of the Mycological Society, Japan* 16: 183-9.
- Yoshida, S. 1985. On the origin of fermented soybeans. In *Non-salted soybean fermentation*. Asian Symposium. International Committee on Economic and Applied Microbiology, 62-3. Tsukuba, Japan.
- Zidwick, M. J. 1992. Organic acids. In *Biotechnology of filamentous fungi*, ed. D. B. Finkelstein and C. Ball, 303-34. Boston, Mass.

II.C.8 🍂 Squash

Definition

Wild and domesticated members of the New World genus *Cucurbita* L. (Cucurbitaceae) are typically referred to as "gourds," "pumpkins," and "squashes." The mature fruit of wild plants, technically called a pepo, has gourdlike qualities like a tough rind and dry flesh. These same qualities have led to the term "ornamental gourds" for various cultivars of *Cucurbita pepo* L. that are grown for their decorative but inedible fruits. However, the common name for the domesticated *Cucurbita ficifolia* Bouché is "fig-leaf gourd," even though the fleshy fruits are cultivated for human consumption. Because another genus of the Cucurbitaceae, *Lagenaria* L., is considered the true gourd, it is preferable to refer to members of *Cucurbita* differentially, which leads us to the terms "pumpkin" and "squash."

Pumpkin comes from the Old English word "pom-pion," which is itself derived from the Greek *pepon* and the Latin *pepo* that together mean a large, ripe,

round melon or gourd. Traditionally, “pumpkin” has been used to describe those cultivars of *Cururbita argyrosperma* Huber, *Cururbita maxima* Lam., *Cururbita moschata* (Lam.) Poir., and *C. pepo* that produce rotund mature fruits used in baking and for feeding livestock.

“Squash,” by contrast, is a term derived from the New England aboriginal word “askutasquash,” meaning vegetables eaten green. It was used during the seventeenth century to designate cultivars, usually of *C. pepo*, grown for their edible immature fruits, and by the nineteenth century, called “summer squashes.” “Winter squashes,” in contrast, are the mature fruits of *C. argyrosperma*, *C. maxima*, *C. moschata*, and *C. pepo* that store well and are not usually round; they are prepared as vegetables, baked into pies, or used as forage. Although “winter squashes” are supposed to differ from “pumpkins” in having a milder taste and flesh of a finer grain, the truth is that these culinary categories overlap, adding to the confusion in nomenclature. For the purposes of this discussion, the generic “squash” will refer to all wild and domesticated members of *Cucurbita*.

Squash Growers and Researchers

The story of squash is a story of Native Americans and New World archaeologists, gold-seeking explorers and European colonizers, herbalists and horticulturists, breeders and botanists. Squashes fascinate us all, but none more than the people who have dedicated their careers to squash research. Such research, as we know it, was under way in Europe by the 1800s. Intrigued by the diversity of fruits illustrated in the herbals of the sixteenth and seventeenth centuries (see Whitaker 1947; Eisendrath 1962; Paris 1989), the French horticulturist Charles Naudin (1856) took great pleasure in describing, breeding, and classifying these newcomers to the Old World. By the twentieth century, comprehensive breeding programs were well established in Europe, North America, and Asia. In an attempt to keep pace with the burgeoning of new strains, William Tapley, Walter Enzie, and Glen Van Eseltine (1937) combed the horticultural literature to provide the most detailed descriptions ever of 132 cultivars.

From Russia, organized plant-collecting expeditions were launched to search Middle and South America, eastern Africa, India, and Asia Minor for new landraces. These explorations provided the bases for new classifications (e.g., Bukasov 1930; Pangalo 1930; Zhiteneva 1930; Filov 1966). Other scientists also contributed to the systematics of squash, with Igor Grebenščikov (1955, 1958, 1969) updating an earlier (Alefeld 1866) classification of infraspecific varieties. The Americans E. F. Castetter and A. T. Erwin took a different approach, placing cultivars into horticultural groups as opposed to botanical classes (Castetter 1925; Castetter and Erwin 1927).

During the middle of the twentieth century, archaeological discoveries of ancient squash in the New World (Whitaker and Bird 1949; Whitaker, Cutler, and MacNeish 1957; Cutler and Whitaker 1961; Whitaker and Cutler 1971) provided an added perspective on the history and evolution of these species. In recent decades, some of the most ancient and most accurately dated and identified squash remains (e.g., Kay, King, and Robinson 1980; Conrad et al. 1984; Simmons 1986; Decker and Newsom 1988) have served to highlight the importance of *C. pepo* in the origins and character of North American horticulture (Heiser 1979; Minnis 1992; Smith 1992). Moreover, archaeological studies in South America have also blossomed recently (see Pearsall 1992 and refs. therein), giving us more detailed histories of *C. ficifolia* and *C. maxima*.

Domesticated squashes, with their diversity in fruit characteristics, have long been of interest to horticultural geneticists (e.g., Sinnott 1922; Shiffriss 1955; Wall 1961; Robinson et al. 1976). Liberty Hyde Bailey, who explored North America in search of wild species, spent countless hours in his gardens performing breeding and inheritance experiments and making observations on the domesticates (Bailey 1902, 1929, 1937, 1943, 1948).

Thomas Whitaker, a prolific researcher with the United States Department of Agriculture, has been the closest human ally of the cucurbits. He examined relationships among wild and domesticated squashes using all available sources of data, including archaeological remains, hybridization experiments, anatomical

Long white squash



and morphological studies, and various genetic analysis (e.g., Whitaker 1931, 1951, 1956, 1968; Whitaker and Bohn 1950; Cutler and Whitaker 1956; Whitaker and Bemis 1964; Whitaker and Cutler 1965). Other devoted squash enthusiasts of the twentieth century include Hugh Cutler and W. P. Bemis, who often worked and published with Whitaker.

In recent years, individual domesticated squash species have been scrutinized to determine their evolutionary histories from wild progenitor(s) through domestication to diversification and geographic spread. As an additional source of phylogenetic data, isozyme analyses aided Deena Decker-Walters and Hugh Wilson in their examination of *C. pepo* (Decker 1985, 1988; Decker and Wilson 1987), Laura Merrick (1990) in the study of *C. argyrosperma*, and Thomas Andres (1990) in his evaluation of *C. ficifolia*. Similar modern and detailed research is lacking for *C. maxima* and *C. moschata*.

Two very different but nonetheless comprehensive books have been written on members of the Cucurbitaceae. One by Whitaker and Glen Davis (1962) reviews past research to provide the most up-to-date (at that time) coverage on the description, history, genetics, physiology, culture, uses, and chemistry of economically important cucurbits, including squashes. The other, *Biology and Utilization of the Cucurbitaceae*, edited by David Bates, Richard Robinson, and Charles Jeffrey (1990), includes 36 distinct articles written by leading experts of the day and covering the systematics, evolution, morphology, sex expression, utilization, crop breeding, and culture of squashes and other cucurbits.

Plant and Fruit Descriptions

Five domesticated and about 20 wild squash species grow in dry or somewhat humid regions of the tropics, subtropics, and mild temperate zones. Their native turf ranges from the central United States south to central Argentina, with species diversity being greatest in Mexico. The herbaceous vines are not frost-tolerant. However, some of the xerophytic perennials have large storage roots that can survive a snowy winter. Among the mesophytic annuals, which include the domesticates, quick germination, early flowering, and rapid growth have enabled some to adapt to the more extreme latitudes.

Squash plants are monoecious, tendrilliferous vines with leaves ranging from entire to lobed and large, yellow to yellow-orange, campanulate flowers. The ephemeral blossoms, opening only once in the morning, are pollinated primarily by specially adapted solitary bees. The inferior ovary of the female flower develops into a gourdlike fruit called a pepo. Pepos of wild plants are usually round with a tough rind and bitter flesh, whereas domesticated fruits generally lack bitterness and are multifarious in their characteristics.

Although primarily outcrossers, individual plants are self-compatible. Hybridization can also occur

between some species. In fact, all squash species have 20 pairs of chromosomes and are incompletely isolated from one another by genetic barriers. This ability to cross species boundaries has been important for plant breeders, allowing them to transfer genes controlling favorable qualities from one species to another. In this way, resistance to the cucumber mosaic virus was transferred from a distantly related wild species to cultivated *C. pepo*, using *C. moschata* as the intermediary.

Archaeological remains, hybridization studies, and genetic data suggest that the domesticated species were independently selected from genetically distinct wild progenitors. In spite of their separate origins, *C. argyrosperma* and *C. moschata* are closely related. In fact, *C. argyrosperma* was not recognized as a distinct species until the Russian geneticist K. I. Pangalo (1930) described it as *Cucurbita mixta* Pang. following extensive collecting expeditions to Mexico and Central America. Even so, it can be difficult to correctly identify some plants and fruits. Generally, fruits of *C. argyrosperma* have enlarged corky peduncles, whereas those of *C. moschata* are hard and thin but distinctly flared at the fruit end. Also, the green and/or white fruits of *C. argyrosperma*, which sometimes mature to yellow, rarely display the orange rind coloring that is common among cultivars of *C. moschata*. Foliaceous sepals are largely unique to but not ubiquitous in *C. moschata*. Leaf mottling is more common in *C. moschata* and leaf lobes deeper in *C. argyrosperma*.

Both species have large flowers with long slender androecia, relatively soft pubescence on the foliage, and distinctly colored seed margins. Among the domesticated species, these squashes best survive the hot, humid, low-elevation (usually under 1,500 meters [m] above sea level) climes of the mid-latitudes, often failing to flower when daylengths are too long. But relative to the wide pre-Columbian distribution and diversity in *C. moschata* (Figure II.C.8.1), *C. argyrosperma* has remained limited in its geography and genetic variability.

There are three domesticated varieties of *C. argyrosperma* subspecies (ssp.) *argyrosperma* - variety (var.) *argyrosperma*, var. *callicarpa* Merrick and Bates, var. *stenosperma* (Pang.) Merrick and Bates - and a weedy variety, var. *palmeri* (Bailey) Merrick and Bates (see Table II.C.8.1). Most of the diversity in this squash can still be found in the endemic landraces of the southwestern United States, Mexico, and Central America. The moderately sized, unfurrowed fruits range from globose to pyriform to long-necked; in the latter, the necks may be straight or curved. Rinds are generally decorated with splotchy, irregular green and white stripes, though in var. *callicarpa*, solid white or green fruits are common and the green coloration is often lacy. Commercial cultivars are few, as culinary quality of the pale yellow to orange flesh is relatively poor in this species. Most of the cultivars and landraces in commercial trade today represent var. *callicarpa*.



Figure II.C.8.1. The white-spotted leaves, yellow-orange female flower, and immature green-and-white-striped fruit of *Cucurbita moschata*. The swelling of the peduncle where it is attached to the fruit is characteristic of this species.

Fruits of *C. moschata*, weighing up to 15 kilograms (kg) apiece, range from squatty to round to turbinate, pyriform, or oblong to necked. Furrows, sometimes deep, are common and wartiness occasional. The rinds are solid, splotchy, or striped in dark to light greens, whites, creams, yellows, and oranges. Fruit flesh is usually deep yellow or orange.

In North America, cultivars of *C. moschata* have been placed into three horticultural groups – “cheese pumpkins,” “crooknecks,” and “bell squashes” (see Table II.C.8.2). However, these groups do not satisfactorily accommodate the diversity of landraces that have evolved in tropical regions around the globe. For example, *C. moschata* is widely cultivated in Asia, and several unusual cultivars with names like ‘Chirimen’, ‘Kikuza’, ‘Saikyo’, and ‘Yokohama’ originated in Japan. Fruit character-

istics resembling those of *C. argyrosperma* ssp. *argyrosperma* indicate that the Japanese cultivars may have arisen from interspecific crossings. Genetic diversity in some northwestern Mexican landraces of *C. moschata* also may be the result of introgression from wild and/or cultivated *C. argyrosperma*.

A. I. Filov (1966) expanded earlier classifications of *C. moschata* to include over 20 varieties in several geographical subspecies. Unfortunately, modern systematic and genetic studies that would confirm the natural relationships among cultivars within and among regions are lacking. Nevertheless, these geographical subspecies do reveal centers of diversification that include Colombia, where the seeds are darker and the plants and fruits small; Mexico, Central America, and the West Indies, where landraces

Table II.C.8.1 Domesticated varieties of *Cucurbita argyrosperma* ssp. *argyrosperma*

Variety	Description	Distribution in North America	Cultivars and comments
var. <i>argyrosperma</i>	Fruits mostly striped; peduncle relatively thin; seeds broad, smooth-surfaced, white with gray margins	Eastern and southern Mexico, Central America	‘Silverseed Gourd’ is grown for its seeds, which are the largest in the genus
var. <i>callicarpa</i>	Fruits solid in color, striped or blotchy; peduncle thick; seeds white or golden tan with tan margins, surfaces smooth or etched	Central and northwestern Mexico, southwestern U.S.	‘Green Striped Cushaw’, ‘Japanese Pie’, ‘Puritan’, ‘Tennessee Sweet Potato’, ‘White Cushaw’, and various landraces; good-quality fruit flesh
var. <i>stenosperma</i>	Fruits mostly striped; dark green tint to the placental tissue; peduncle thick; seeds narrow, smooth-surfaced, white with gray or tan margins	South-central Mexico	‘Elfrida Taos’; distribution and characteristics overlap with those of the other varieties; grown mostly for the edible seeds

Table II.C.8.2. Horticultural groups of *Cucurbita moschata*

Horticultural group	Description	Representative cultivars	Comments
Cheese pumpkins	Fruits variable but usually oblate with a buff-colored rind	'Calhoun', 'Kentucky Field', 'Large Cheese', 'Quaker Pie'	Plants are hardy and productive under various growing conditions
Crooknecks	Fruits round at blossom end with long straight or curved necks	'Bugle Gramma', 'Canada Crookneck', 'Golden Cushaw', 'Winter Crookneck'	Very popular in colonial America for pies and stock
Bell squashes	Fruits bell-shaped to almost cylindrical	'African Bell', 'Butternut', 'Carpet Bag', 'Ponca', 'Tahitian'	These cultivars, which are the most popular today, were probably selected from "crookneck" types

are genetically variable and fruits of many shapes and colors can be found in a single field; Florida, which is home to the small-fruited, aboriginal 'Seminole Pumpkin'; Japan with its warty and wrinkled fruits; India, where large soft-skinned fruits abound; and Asia Minor, where fruits again are variable but long barbell-shaped pepos predominate.



Figure II.C.8.2. Seeds of *Cucurbita pepo* (top), *C. moschata* (center), and *C. argyrosperma* (bottom). At the lower right is a seed of 'Silverseed Gourd,' the largest of all squash seeds. Scale equals 1 cm.

Although *C. moschata* is the most widely cultivated squash in underdeveloped tropical countries, as with *C. argyrosperma*, relatively few cultivars have entered the commercial trade of Europe and North America. "Cheese pumpkins" and "crooknecks" were popular in nineteenth-century New England. Today, only various selections from the "bell squashes" are commonly sold by seed suppliers (Figure II.C.8.3).

Cucurbita pepo is characterized by uniformly colored tan seeds, lobed leaves with prickly pubescence, hard roughly angled peduncles, and short, thick, conical androecia (Figures II.C.8.2 and II.C.8.4). Flowers range from small to large, though they are rarely as grand as those of *C. argyrosperma* ssp. *argyrosperma*. Genetic diversity, expressed in the plethora of differing fruit forms, is greatest in this squash. Orange flesh color is not as common in *C. pepo* as it is in *C. maxima* or *C. moschata*.

Cucurbita pepo appears to have shared a Mexican or Central American ancestor with *C. argyrosperma* and *C. moschata*. From those origins, wild populations – ssp. *ovifera* (L.) Decker var. *ozarkana* Decker-Walters, ssp. *ovifera* var. *texana* (Scheele) Decker, ssp. *fraterna* (Bailey) Andres, ssp. *pepo* – spread over North America before at least two domestications of *C. pepo* took place to produce ssp. *ovifera* var. *ovifera* (L.) Decker and cultivars of ssp. *pepo*. Because *C. pepo* can tolerate cooler temperatures than can *C. argyrosperma* and *C. moschata*, this squash flourishes at more extreme latitudes and higher elevations (1,600 to 2,100 m above sea level) to the delight of farmers from southern Canada to the highlands of Central America. Some wild populations and cultivars are well adapted to the northern United States, with seeds that are quick to germinate and early flowering that is responsive to changes in daylength.

Encompassing many hundreds of cultivars, six horticultural groups of *C. pepo* were recognized during the twentieth century (see Table II.C.8.3). "Acorn squashes," "crooknecks," "scallop squashes," and most "ornamental gourds" belong to ssp. *ovifera* var. *ovifera*. Horticulturists have traditionally classified all small, hard-shelled, bitter fruits grown



Figure II.C.8.3. 'Butternut', a "bell squash" cultivar of *Cucurbita moschata*, has a cream-colored rind and dark orange flesh rich in carotenes.

for autumn decorations as ornamental gourds. However, this classification does not reflect the fact that these gourds have various genealogies that include wild populations of ssp. *ovifera*, ssp. *pepo*, and probably ssp. *fraterna*. Pumpkins, such as those grown in temperate to tropical gardens around the globe, and marrows belong to ssp. *pepo*. The former, like acorn squashes, are eaten when mature, whereas the latter, like the crooknecks and scallop squashes, are summer squashes picked during the first week of fruit development. Bushy plants with relatively short internodes have been developed for most of the summer squashes as well as for some of the acorn squashes (Figures II.C.8.5 and II.C.8.6).

Cucurbita maxima is distantly related to the trio just discussed. This squash, whose origins are in South America, exhibits closer genetic affinities to other wild South American species. Like *C. pepo*, some cultivars and landraces of *C. maxima* can tolerate the relatively cool temperatures of the highlands (up to 2,000 m above sea level). Today, this species is grown in tropical to temperate regions around the globe, particularly in South America, southeastern Asia, India, and Africa.

Cucurbita maxima is distinguished by its soft round stems, entire or shallowly lobed unpointed



Figure II.C.8.4. An unusual "acorn squash" of *Cucurbita pepo* purchased in Ontario, Canada; cultivar unknown. Flesh and rind are tan colored.

leaves, and spongy, enlarged, terete peduncles. Compared to the other domesticates, the white or brown seeds of this squash are thick, particularly in relationship to their margins. The androecium is short, thick, and columnar. The yellow or orange fruit flesh is fine-grained and of the highest quality (tasty and relatively rich in vitamins) among all squashes. Fruits are quite variable in size, shape, and coloration, with the latter including many shades of gray, green, blue, pink, red, and orange in striped, mottled, or blotchy patterns.

A distinct fruit form characterizes cultivars classified as "turban squashes" (Figures II.C.8.7 and II.C.8.8). The immature ovary protrudes upward from the receptacle, swelling into a turban-shaped fruit with a crown (the part of the fruit not enveloped by the receptacle) upon maturity. Table II.C.8.4 lists some turban squash cultivars and describes five other traditionally recognized horticultural groups – "banana squashes," "delicious squashes," "hubbard squashes" (Figure II.C.8.9) "marrows," and "show pumpkins."

Table II.C.8.3. *Horticultural groups of Cucurbita pepo*

Horticultural group	Description	Representative cultivars	Comments
Acorn squashes	Fruits usually small, of various shapes and colors but always grooved	'Acorn', 'Delicata', 'Fordhook', 'Mandan', 'Sweet Dumpling', 'Table Queen'	A heterogeneous group of uncertain origins but closely related to "scallop squashes"; mature fruits baked as vegetable
Crooknecks	Fruits long, club-shaped with straight or curved neck; rind very hard, yellow to orange, warted	'Giant Crookneck', 'Straightneck', 'Summer Crookneck', 'Yankee Hybrid'	Probably an ancient group of summer squashes although 'Straightneck' cultivars are more recent in origin
Marrows	Fruits long, club-shaped to cylindrical, mildly ridged; rind usually with lacy green pattern	'Cocozelle', 'Moore's Cream', 'Vegetable Marrow', 'Zucchini'	Selected from pumpkins and diversified in Europe; fruits eaten immature
Ornamental gourds	Fruits small of various shapes and colors; rind hard; flesh usually bitter; seeds small	'Crown of Thorns', 'Flat Striped', 'Miniature Ball', 'Nest Egg', 'Orange Warted', 'Spoon', 'Striped Pear'	A heterogeneous group of multiple origins; some cultivars primitive, others fairly new; fruits not eaten
Pumpkins	Fruits typically large, round or oblong, shallowly to deeply grooved or ribbed; rind relatively soft; seeds large	'Connecticut Field', 'Jack O'Lantern', 'Sandwich Island', 'Small Sugar', 'Vegetable Spaghetti'	Mature fruits used as a vegetable or for pies, jack-o'-lanterns, and forage; grown for edible seeds also
Scallop squashes	Fruits flattened at both ends with edges scalloped around middle; rind hard	'Benning's Green Tint', 'Golden Custard', 'Patty-pan', 'White Bush Scallop'	An ancient group of summer squashes

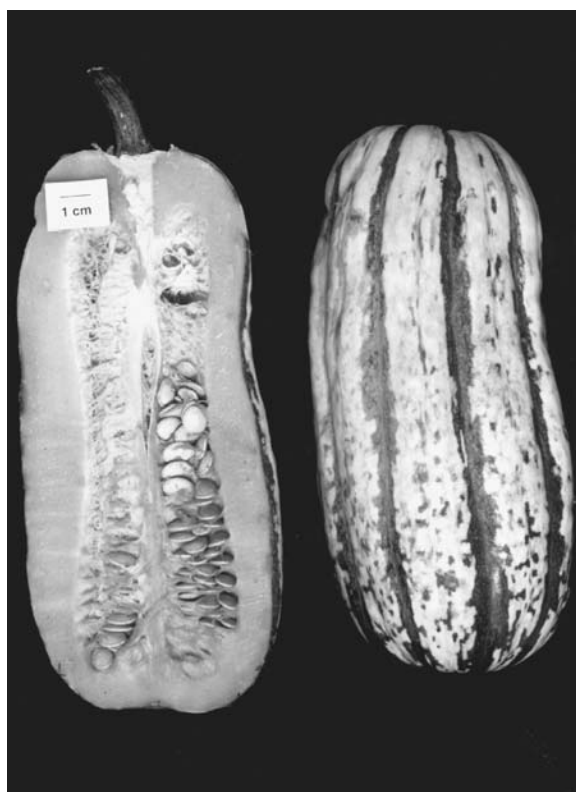


Figure II.C.8.5. 'Delicata' (*Cucurbita pepo*) has a green-and-white-striped rind that turns orange and pale yellow with age. The orange flesh of these long-keeping fruits intensifies in color with storage.



Figure II.C.8.6. Various small-fruited cultivars of *Cucurbita pepo*, including 'Sweet Dumpling' (top left), 'Bicolor Spoon' (top right), 'Orange Warted' (center), 'Table Queen' (bottom left), and 'Jack-Be-Little Pumpkin' (bottom right).



Figure II.C.8.7. 'Turk's Turban' (*Cucurbita maxima*) is often grown as an ornamental because of the deep red rind color. The crown of this fruit has a much paler pattern of green and red splotches.

Over 50 cultivars of *C. maxima* had been commercially traded by the early twentieth century; today, this number has reached over 200. Not all landraces and cultivars can be assigned to the horticultural groups in Table II.C.8.4. Some cultivars, such as the warty 'Victor', were derived from hybridizations between

horticultural groups. Local landraces that never entered into, or played only minor roles in, American and European commercial trade often have fruit traits that do not match those characterizing the groups. And although several varieties of *C. maxima* have been proposed over the years, as of yet no one has performed an intensive systematic study of this species to clarify evolutionary relationships among cultivars and groups of cultivars.

Cucurbita ficifolia is not closely related to the other domesticated squashes or to any known wild populations. Distinguishing characteristics include relatively wide, minutely pitted, solid-colored seeds, ranging from tan to black; white, coarsely fibrous flesh; an androecium shaped like that of *C. maxima* but with hairs on the filaments; and rounded lobed leaves. Genetic uniformity in this species is evidenced by the lack of variation in fruit characteristics. The large oblong fruits, measuring 15 to 50 centimeters (cm) long, exhibit only three basic rind coloration patterns: solid white, a reticulated pattern of green on white that may include white stripes, and mostly green with or without white stripes. No distinct landraces or cultivars of *C. ficifolia* have been recognized.

In Latin America today, this cool-tolerant, short-day squash is grown for food in small, high-altitude (1,000 to 2,800 m above sea level) gardens from northern Mexico through Central America and the Andes to central Chile. Usually the mature fruits are candied, but the seeds and immature fruits are eaten as well. *Cucurbita ficifolia* is also cultivated as an ornamental

Table II.C.8.4. Horticultural groups of *Cucurbita maxima*

Horticultural group	Description	Representative cultivars	Comments
Banana squashes	Fruits long, pointed at both ends; rind soft; seeds brown	'Banana', 'Pink Banana', 'Plymouth Rock'	Introduced to the U.S. from Mexico; plants can tolerate high temperatures
Delicious squashes	Fruits turbinate, shallowly ribbed; rind hard; seeds white	'Delicious', 'Faxon', 'Quality'	High-quality flesh; original stock came from Brazil in the late 1800s; similar types occur in Bolivia
Hubbard squashes	Fruits oval, tapering curved necks at both ends; rind very hard; seeds white	'Arikara', 'Blue Hubbard', 'Brighton', 'Hubbard', 'Kitchenette', 'Marblehead'	Inbreeding and hybridization have produced many cultivars
Marrows	Fruits oval to pyriform, tapering quickly at the apex and gradually toward the base; seeds white	'Boston Marrow', 'Golden Bronze', 'Ohio', 'Valparaiso', 'Wilder'	Fruits mature early; original stock probably came from Chile
Show pumpkins	Fruits large, orange; rind soft; seeds white	'Atlantic Giant', 'Big Max', 'Big Moon', 'Etampes', 'Mammoth'	Produces the largest pepos in the genus; grown for show and forage; a lot of diversity in India
Turban squashes	Fruits turban-shaped	'Acorn', 'Buttercup', 'Crown', 'Essex', 'Red China', 'Sweetmeat', 'Turban', 'Warren', 'Sapallito del Tronco'	Many cultivars were selected in Africa, Asia, and Australia from South American stock; some are more bushy than viny

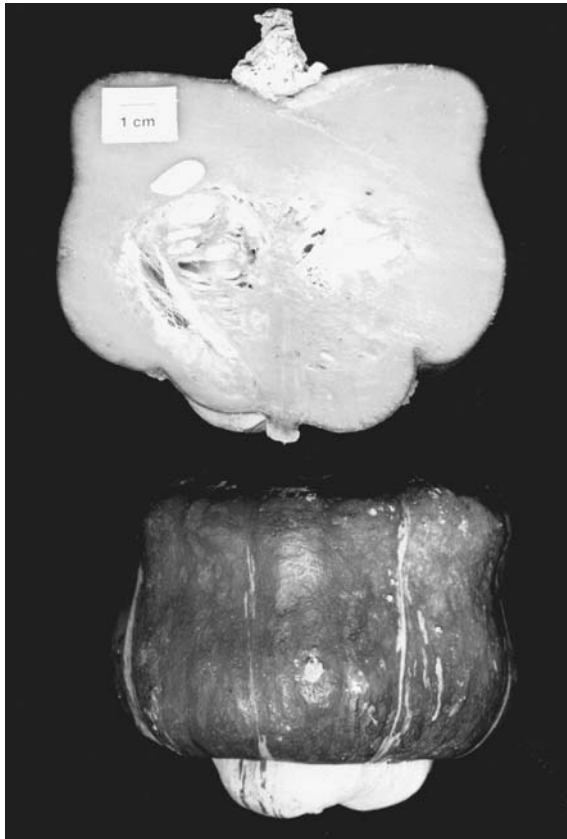


Figure II.C.8.8. 'Buttercup' is a "turban squash" of *Cucurbita maxima*. This fruit has a dark green rind, except for the small crown, which is a pale blue-green. Seeds are white and the flesh dark orange.

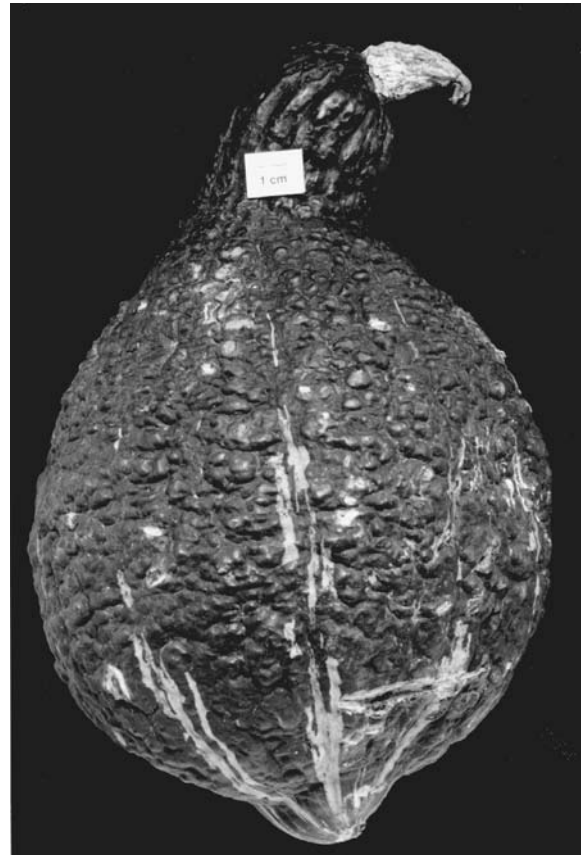


Figure II.C.8.9. This "hubbard squash" of *Cucurbita maxima* has a warty, dark blue-green rind with a few pale stripes. Note the swollen peduncle.

in Europe and the United States and for forage in underdeveloped countries of the Old World.

The Evolution and History of Squashes

The five domesticated squash species were brought under cultivation 5,000 to 15,000 years ago. Native Americans transformed the small green and white gourdlike pepos of wild plants into a cornucopia of colorful and shapely pumpkins and squashes. But long before they were domesticated, wild squash plants made their presence known to human populations. These fast-growing, tenacious vines are prolific opportunists, boldly invading disturbed sites of natural or human origin. Initial human interest in wild squash may have manifested itself in the use of the tough fruit rinds for containers. Additionally, the seeds are a tasty and nutritious source of food. The flesh of wild pepos is too bitter to eat raw. Toxic oxygenated tetracyclic triterpenes, called cucurbitacins, permeate the leaves, roots, and fruits as deterrents to

herbivory. Nevertheless, the frequency of immature peduncles among archaeological remains suggests that the young tender fruits were consumed, probably after leaching out the cucurbitacins through multiple boilings.

The development of nonbitter pepos came about as a result of domestication. Indiscriminate harvesting of fruits from wild or tolerated weedy vines eventually led to the planting of seeds from selected genetic strains. In the process of selecting for larger fruits for containers or larger seeds for consumption, thicker, nonbitter, and less fibrous flesh was discovered and selected for as well. Other changes included the loss of seed dormancy, softer and more colorful fruit rinds, adaptation to shorter growing seasons, and generally larger plant parts. In this way, squash became a major component of diets for the ancient farmers of the New World.

Squash domestication took place at least five times to yield the cultivated members of *C. argyrosperma*, *C. ficifolia*, *C. maxima*, *C. moschata*, and *C. pepo*.

These domestications involved genetically distinct wild populations and at least three different cultural groups inhabiting the eastern United States, Middle America, and South America. A discussion of the evolution of these cultivated squashes along with their history and spread follows.

Cucurbita argyrosperma

Cultivars of *C. argyrosperma* ssp. *argyrosperma* are genetically similar to wild populations of *C. argyrosperma* ssp. *sororia*, a native of low-elevation, mostly coastal habitats in Mexico and Central America. Sufficient evidence exists to support the theory that ssp. *sororia* gave rise to domesticated ssp. *argyrosperma*. Domestication probably took place in southern Mexico, where the earliest remains of ssp. *argyrosperma* date around 5000 B.C. Most of these archaeological specimens belong to var. *stenosperma*; landraces of this variety are still grown in southern Mexico today. With a current distribution ranging from northeastern Mexico south to the Yucatan and into Central America, var. *argyrosperma* is the most widespread variety of ssp. *argyrosperma* (Figure II.C.8.10). Remains of var. *argyrosperma* first appear in the archaeological

record in northeastern Mexico at about A.D. 200. A little later (c. A.D. 400), var. *callicarpa* shows up at archaeological sites in the southwestern United States. The earliest pre-Columbian evidence of *C. argyrosperma* in eastern North America is fifteenth-century remains from northwestern Arkansas. Although the three varieties of ssp. *argyrosperma* could have been selected separately from wild populations of ssp. *sororia*, Merrick's (1990) interpretation of the morphological evidence suggests that var. *stenosperma* and var. *callicarpa* evolved from southern and northern landraces of var. *argyrosperma*, respectively.

The fourth and final variety of ssp. *argyrosperma*, var. *palmeri*, is weedy, possessing a mixture of characteristics otherwise representing wild ssp. *sororia* and cultivated var. *callicarpa*. It grows unaided in disturbed areas, including cultivated fields, in northwestern Mexico beyond the range of ssp. *sororia*. *Cucurbita argyrosperma* ssp. *argyrosperma* var. *palmeri* may represent escapes of var. *callicarpa* that have persisted in the wild by gaining through mutation and/or hybridization with ssp. *sororia* those characteristics (such as bitter fruits) that are necessary for independent survival.

Current uses of wild and weedy fruits in Mexico include eating the seeds, using seeds and the bitter flesh medicinally, washing clothes with the saponin-rich flesh, and fashioning containers from the dried rinds. Although the antiquity of these uses is uncertain, selection for edible seeds has remained the dominant theme in cultivation. In southern Mexico and Guatemala, var. *argyrosperma* and var. *stenosperma* are grown for their large edible seeds, with the fruit flesh serving as forage. In southern Central America, indigenous cultures have produced landraces that yield a necked fruit eaten as a vegetable while immature and tender. Selection pressures in northern Mexico have created several landraces of var. *argyrosperma* and var. *callicarpa*; some produce mature pepos with quality flesh for human consumption as well as edible seeds, whereas others are grown for their immature fruits. At twelfth- and thirteenth-century sites in southern Utah, fruits of var. *callicarpa* were employed as containers, a use that persists among some tribes of the Southwest today. The greatest diversity of fruits in the species, represented primarily by var. *callicarpa*, occurs in northwestern Mexico and the southwestern United States.

Relative to the post-Columbian changes incurred by other squashes, the spread and further diversification of *C. argyrosperma* cultivars has been limited. A few commercial cultivars such as 'Green Striped Cushaw' were selected from North American stock and grown in New England soon after colonization. A similar type of squash was illustrated in European herbals of the sixteenth century, and additional types were developed in South America and Asia. As a result of the recent trend to identify, save, and distribute native landraces of New World crops, a large number

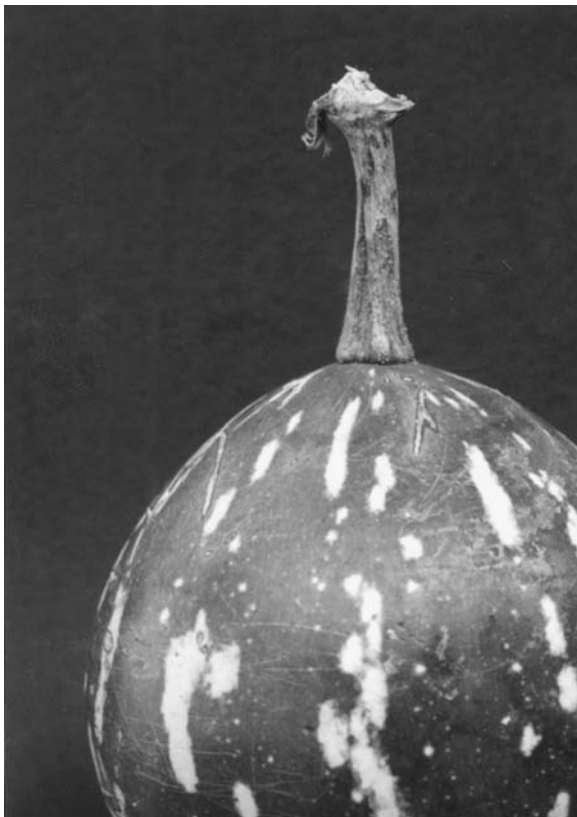


Figure II.C.8.10. This mature fruit of *Cucurbita argyrosperma* ssp. *sororia*, measuring about 7 cm. in diameter, was collected from a wild population in Veracruz, Mexico.

of landraces of *C. argyrosperma* indigenous to North America have entered the U.S. commercial trade under such names as 'Chompa', 'Green Hopi Squash', 'Mayo Arrote', 'Montezuma Giant', and 'Pepinas'.

Cucurbita moschata

The earliest archaeological remains indicative of domestication of *C. moschata* were discovered in southern Mexico (5000 B.C.) and in coastal Peru (3000 B.C.). Ancient Peruvian specimens differ from those of Mexico by having a warty rind and a pronounced fringe along the seed margins. Although Mexico appears to be the more ancient site of domestication, the Peruvian remains and the diversity of Colombian landraces point to South America as a secondary site of early diversification or an independent center of domestication. Unfortunately, wild progenitor populations have not yet been identified for this species. It is possible that they are extinct; however, a few tantalizing finds of wild squash in Bolivia suggest that South American populations of *C. moschata* may be awaiting rediscovery. Among wild squashes known today in Middle America, those of *C. argyrosperma* ssp. *sororia* express the greatest genetic affinity to Mexican landraces of *C. moschata*.

Even though the centers of landrace diversity for *C. moschata* lie in Central America and northern South America, archaeological remains indicate that this species spread to northeastern Mexico by about 1400 B.C. and to the southwestern United States by A.D. 900. The spread of *C. moschata* to the Gulf coastal area and the Caribbean may have been facilitated by early Spanish explorers and missionaries; a distinctive Florida landrace called 'Seminole Pumpkin' (Figure II.C.8.11) is still grown by the Miccosukees of the Everglades.

Among tribes of the Northern Plains, *C. moschata* was definitely a post-Columbian introduction.

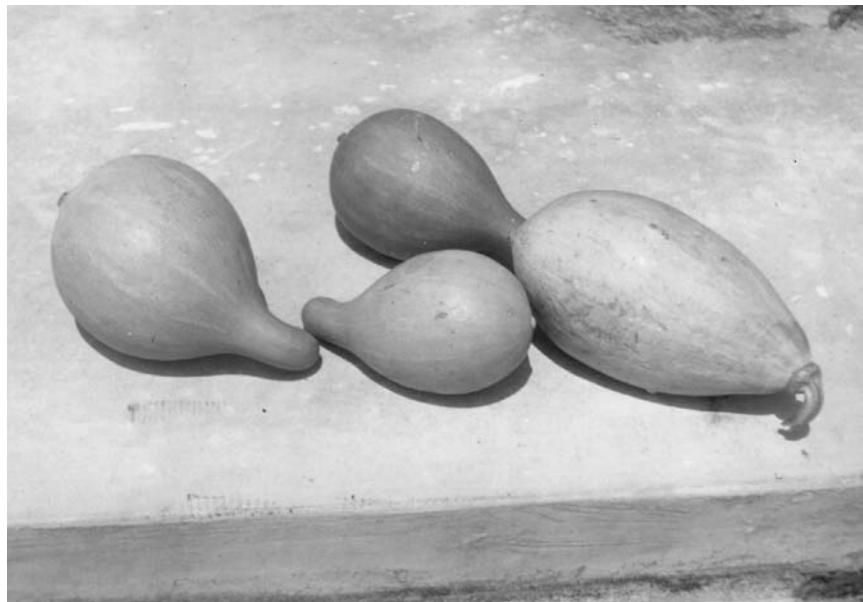
The crooknecks and cheese pumpkins, which probably have their origins in North America, were known to colonists and Europeans as early as the 1600s. Variations on the cheese pumpkin theme can be found in the large furrowed pumpkins of India and southeastern Asia. Additional diversification of *C. moschata* took place in Asia Minor, where various fruit types resemble the bell squashes, and in Japan, where selection was for heavily warted rinds. Completing its worldwide travels, this species was well established as a food crop in northern Africa by the nineteenth century.

Cucurbita pepo

The squash represented by the earliest archaeological remains is *C. pepo*. Its seeds and rinds of wild or cultivated material appear in Florida around 10,000 B.C., in southern Mexico around 8000 B.C., and in Illinois at around 5000 B.C. Enlarged seeds and peduncles as well as thicker fruit rinds suggest that this species was definitely under cultivation in southern and northeastern Mexico between 7500 and 5500 B.C. and in the Mississippi Valley between 3000 and 1000 B.C. Cultivation had spread from Mexico to the southwestern United States by around 1000 B.C., and by A.D. 1500, *C. pepo* landraces were being grown throughout the United States and Mexico.

Ancestral North Americans independently domesticated at least two genetically distinct and geographically disjunct subspecies of *C. pepo* to produce the two major lineages of cultivars known today. Although wild populations of ssp. *pepo* (Figures II.C.8.12 and II.C.8.13) are currently unknown

Figure II.C.8.11. Fruit shape in the tan-colored fruits of the 'Seminole Pumpkin' (*Cucurbita moschata*) ranges from round or oblate to oval, oblong, or pear shaped. (Photograph by John Popenoe.)



and possibly extinct, they were probably subjected to the selection pressures of the natives of southern Mexico, giving rise to the majority of Mexican and southwestern U.S. landraces, as well as to “pumpkin” and “marrow” cultivars of this species. As with *C. argyrosperma* and *C. moschata*, human selection with *C. pepo* landraces in southern Mexico focused on producing large seeds within a sturdy round fruit.

Today, wild populations of *C. pepo* range from northeastern Mexico (ssp. *fraterna*) to Texas (ssp. *ovifera* var. *texana*), and north through the Mississippi Valley to southern Illinois (ssp. *ovifera* var. *ozarkana*).

As recently as 250 years ago, wild populations of ssp. *ovifera* may have occurred throughout the Gulf coastal region and certainly in Florida. A whole different lineage of cultivars, classified as ssp. *ovifera* var. *ovifera*, evolved from eastern U.S. populations of var. *ozarkana*. Aborigines of the Mississippi Valley apparently were not as interested as the Mexicans in quickly selecting for large seeds or fleshy fruits. Instead, a variety of small, odd-shaped, hard, and often warty cultivars were used as containers or grown for other nonsubsistence purposes. And although the seeds of early cultivars were probably eaten, in selecting for food, natives of the eastern United States developed several cultivars, such as the precursors of the scallop squashes and crooknecks, that produced tasty immature fruits.

Gilbert Wilson's (1917) treatise on agriculture among the Hidatsa indigenes of the Northern Plains gives us a more detailed account of the aboriginal use of *C. pepo*. These Native Americans cultivated squashes of various shapes, sizes, and colors together, picking them when four days old.



Figure II.C.8.12. The white fruit of this wild *Cucurbita pepo* ssp. *ovifera* var. *ozarkana* plant still clings to the dead vine at a riparian site in the Mississippi Valley. (Photograph by Wes Cowan and Bruce Smith.)



Figure II.C.8.13. From left to right, these green-and-white-striped fruits of *Curcubita pepo* represent wild ssp. *ovifera* var. *texana*, ‘Mandan’ (a cultivar of var. *ovifera*), and wild ssp. *fraterna*.

The young fruits were eaten fresh or sliced and dried for the winter. The flesh, and sometimes the seeds, of these mature fruits were boiled and eaten. Male squash blossoms did not go to waste either; they were picked when fresh and boiled with fat or dried for later use in mush. One fruit per plant was allowed to mature so as to provide seed for the next planting.

In addition to the two primary centers of domestication of *C. pepo*, a few landraces and cultivars may have been domesticated from wild populations of *ssp. fraterna* in northeastern Mexico. These landraces and those from southern Mexico probably spread to the eastern United States between A.D. 1000 and 1500, if not earlier. The intermingling of cultivars undoubtedly gave rise to new genetic combinations, which accounts for the diversity of fruits encountered by the earliest European visitors. The "acorn squashes," which include the Northern Plains landrace 'Mandan', may have originated in this way when Mexican "pumpkins" met "scallop squashes" in the United States. Similarly, 'Fort Berthold' and 'Omaha' are northern-adapted "pumpkin" landraces that were being grown by Sioux tribes in the early twentieth century.

Fruits representing all of the major horticultural groups are pictured in the herbals of the sixteenth century. More than any other squash, *C. pepo* was enthusiastically embraced by European horticulturists; hundreds of new cultivars, particularly the "marrows," have been developed in Europe and in the United States over the past 400 years. Selection practices emphasized earliness in flowering, compactness or bushiness in growth, and uniformity within a cultivar for fruit characteristics.

Although *C. pepo* was carried to other parts of the globe during the seventeenth century, diversification of landraces was limited primarily to the "pumpkins" of Asia Minor. Nevertheless, unique cultivars did develop elsewhere, such as 'Alexandria' from Egypt, 'Der Wing' from China, 'Nantucket' from the Azores, and 'Pineapple' from South America.

Cucurbita maxima

Numerous landraces of *C. maxima* with differing fruit characteristics can be found throughout South America today. However, archaeological remains are less widespread. Most are from coastal Peru, where the earliest evidence of domestication appears between 2500 and 1500 B.C. Later pre-Columbian remains have been found in Argentina (500 B.C.) and northern Chile (A.D. 600). Early Spaniards noted that landraces of *C. maxima* were being grown by the Guaraní indigenes of northeastern Argentina and Paraguay.

The wild progenitor of domesticated *C. maxima* *ssp. maxima* is *C. maxima ssp. andreana* (Naud.) Filov, a weedy native of warm temperate regions in northern Argentina, Uruguay, Bolivia, and possibly

Paraguay. Hybridization between cultivars and wild populations has contributed to genetic variability in *ssp. andreana*, producing wild fruits that vary from pear-shaped to oblong to round. Some landraces may have been selected from these introgressed populations.

South American aborigines apparently selected for large fruits of *ssp. maxima* with high-quality flesh and good storage capabilities. The largest South American fruits, weighing 20 to 40 kg, are found in landraces from central Chile. Fruits with a woody skin suitable for long storage were noted in Bolivia by Russian explorers in the 1920s. Warty fruits also evolved in South America, and in the twentieth century are found in Bolivia and Peru. Other native landraces yield tasty immature fruits.

Cultivation of *C. maxima* did not spread to northern South America, Central America, and North America until after the European invasion of the sixteenth century. Yankee sailors were supposedly responsible for introducing various cultivars, including 'Acorn', or 'French Turban', 'Cocoa-Nut', and 'Valparaiso', to New England early in the nineteenth century. Although all of the horticultural groups probably have their origins in South America, the spread of *C. maxima* throughout North America yielded some new landraces and cultivars. For example, 'Arikara' and 'Winnebago' are landraces that were grown by aboriginal tribes in North Dakota and Nebraska, respectively, during the beginning of the twentieth century. The "banana squashes" proliferated in Mexico, and "Hubbard squashes," like 'Marblehead', came to the eastern United States from the West Indies during the 1800s.

Visitors took several types of *C. maxima* back to Europe during the sixteenth through nineteenth centuries. Some, like the turban squash called 'Zapallito del Tronco' or 'Tree Squash', came directly from South America. Most cultivars were introduced from colonial North America, but others reached Europe via Asia, Australia, and Africa, where local landraces evolved. For example, 'Red China' is a small turban squash that was brought to Europe from China in 1885. India also became a secondary center of cultivar diversity, particularly for the large "show pumpkin" types. Today, Indian fruits, weighing up to 130 kg, range from spherical to oblong to turban-shaped. Unusually shaped squashes with brown seeds such as 'Crown', 'Triangle', and 'Queensland Blue' are late-maturing Australian cultivars. And in Africa, *C. maxima* was so widespread by the nineteenth century that some botanists mistakenly concluded that Africa was the ancestral home of this squash.

In addition to collecting cultivars from around the globe, the Europeans succeeded in producing their own new strains, particularly in nineteenth-century France. For example, 'Etampes' and 'Gray Boulogne' entered the commercial trade in the 1880s. Selections within the "turban squashes" at this time focused on producing smaller nonprotruding crowns.

Cucurbita ficifolia

Pre-Columbian remnants of domesticated *C. ficifolia* have been found only in Peru, with the earliest seeds, peduncles, and rind fragments dating between 3000 and 6000 B.C. An archaeological seed from southern Mexico that was originally identified as *C. ficifolia* apparently belongs to *C. pepo* instead (cf. Andres 1990). Assuming domestication in northern South America, it is not known when this squash reached Mexico; however, it was being cultivated there in the twelfth century.

No wild species of squash exhibits the type of relationship with *C. ficifolia* that is expected in the pairing of a crop with its wild progenitor. Although definitively wild populations of *C. ficifolia* have not been identified, reports of weedy, self-sustaining plants in Guatemala and Bolivia are intriguing and need to be explored further.

As with the other domesticates, human selection produced relatively large, fleshy, nonbitter fruits with large seeds. However, the overall lack of genetic diversity in *C. ficifolia* relative to the other domesticated squashes suggests that human selection pressures on the former have been limited in their duration, their intensity, their diversity, and their effects.

This cool-tolerant but short-day squash did not reach Europe until the early 1800s, coming by way of southern Asia, where the long-keeping fruits were mainly used to feed livestock, especially during lengthy sea voyages. Although some accessions of *C. ficifolia* have been successfully grown as far north as Norway, the general failure of this species to flower beyond the torrid zone may account in part for its lack of popularity in Europe.

Squash Preparation and Consumption

In rural gardens around the globe, squash is typically planted among other crops, particularly corn, and the vines are allowed to scramble over poles, fences, walls, and other nearby structures. The plants like fertile aerated soil that drains well and lots of space, water, and sunshine. Extremes in temperature or high humidity increase vulnerability to disease. During wet weather, placing a stone or fibrous mat under a fruit lying on the ground prevents the fruit from rotting.

The immature and mature fruits, seeds, flowers, buds, and tender shoot tips and leaves of all of the domesticated squashes can be and have been eaten. Harvest of the one- to seven-day-old fruits of the summer squashes begins seven to eight weeks after planting and continues throughout the growing season. Pumpkin and winter squash fruits take three to four months to mature and are harvested only once, usually along with or later than other field crops. The best seeds are taken from the oldest fruits. Once flowering begins, open male blossoms can be collected almost daily. Leaves and growing tips are picked when needed, but only from healthy, vigorous plants.

Even though immature squashes can be eaten raw,

usually they are boiled first and then seasoned to taste. In various cultures, the fresh fruits are sliced, battered, and fried; stuffed with cooked meat and vegetables; boiled and then mashed like potatoes; or added to curries or soups. The Sioux of the Northern Plains sliced fresh four-day-old fruits of *C. pepo*, skewered the slices on willow spits, and placed the spits on open wooden stages for drying. In Mexico and Bolivia, young fruits and seeds of *C. ficifolia* are sometimes blended into a mildly sweetened, alcoholic beverage made from corn mush.

The precursor of the colonial pumpkin pie was a mature pumpkin, probably *C. pepo*, filled with fruit, sugar, spices, and milk. Seeds were removed and ingredients added through a hole cut in the top of the pumpkin. The stuffed fruit was then baked among the coals of an open fire. In a simpler version of this recipe, prepared by aborigines as well as settlers, the fruits were baked first and then sliced and garnished with animal fat and/or honey or syrup. Pumpkin pudding, pancakes, bread, butter, dried chips, and beer have long histories rooted in colonial New England.

In other countries, mature pumpkins and winter squashes are stewed or steamed as vegetables, added to soups, candied, or stored whole or in slices for later use. A presumably ancient aboriginal use of mature fruits of *C. ficifolia* is in making various types of candy. Chunks of flesh are boiled with the seeds in alkali and then saturated with liquid sugar. In Indonesia, the local inhabitants create a delicacy by adding grated coconut to the boiled flesh of *C. moschata*. Of course, the most popular nonfood usage of pumpkin fruits (usually *C. pepo* or *C. maxima*) is for carving jack-o'-lanterns, a nineteenth-century tradition from Ireland and Great Britain.

Although the fruits of all domesticated squashes can be prepared similarly, there are culinary differences among the species with respect to the flavor, consistency, and appearance of the edible flesh. *Cucurbita moschata* and *C. maxima* produce the strongest tasting (mildly sweet and somewhat musky) and deepest colored mature fruits; consequently, these species are favored for canning. Because fruits of *C. maxima* are also the richest in vitamins and finest in texture, they are mashed into baby food. Among the squashes, flesh quality in *C. maxima* generally holds up best when dehydrated and then reconstituted. The elongated fruits of summer squashes make *C. pepo* the foremost producer of easy-to-slice young fruits. Although this species dominates the commercial market, the fuller flavor of the immature pepos of *C. moschata* make *C. moschata* the preferred vegetable in rural China, the Canary Islands, and Central America.

Landraces of *C. argyrosperma* yield the largest edible seeds in a fruit that is otherwise unremarkable. Mature fruits of *C. ficifolia* are the most bland and fibrous of all squashes. However, they store longer than the fruits of the other species (two to three years versus one year) and sweeten with age. The

flesh contains a proteolytic enzyme that may be of future commercial value to the food industry. Because of the stringiness of the flesh of *C. ficifolia*, a special Aztec confection called "Angel's Hair" can be prepared from the boiled flesh fibers. Comparable texture in the *C. pepo* cultivar 'Vegetable Spaghetti' allows preparation of the baked or boiled fibrous flesh into a dish resembling the namesake pasta.

For commercial canning, growers have selected high-yielding cultivars like 'Kentucky Field' that have mature fruit flesh of the proper color and consistency. Flavor is less important as it can be controlled with spices. Consistency, which refers to the stiffness or relative viscosity of the processed flesh, is enhanced by using fruits that are barely ripe and by adding the product of a high-consistency cultivar to that of a low-consistency cultivar. Starch, as well as soluble solids, greatly influences consistency. Because fruit storage results in the loss of carbohydrates and in the conversion of starch to sugars, freshly harvested fruits are preferred for the canning process.

Squash seeds, which have a nutty flavor, are eaten worldwide. They are consumed raw, boiled, or roasted, usually with the testa or shell removed. Mexicans grind the roasted shelled seeds into a meal, which is used to make special sauces. In China and India as well as in the New World, rural peoples make pastries from the seeds, often by covering them with syrup and then baking the mass into a type of peanut brittle. Some Chinese cultivars of *C. moschata* and *C. maxima* are grown specifically for their seeds. Similarly, various landraces of *C. argyrosperma* contribute heavily to the commercial production of edible seeds in Mexico.

A "naked seed" cultivar of *C. pepo*, called 'Lady Godiva', produces a seed lacking a testa. These hull-less seeds are popular snacks in the United States. In addition to food, New World aborigines have used squash seeds for a variety of medicinal purposes. A decoction

serves as a diuretic and an antipyretic, the seed oil is applied to persistent ulcers, and the seeds are eaten to expel gastrointestinal parasites. Although rural communities use the seed oil for cooking as well as for medicine, the possibility of commercial extraction of the edible unsaturated oil has yet to be explored.

Aboriginal Americans, including the Aztecs, have a long tradition of eating male squash flowers and floral buds. The large orange blossoms lend seasoning and color to stews, soups, and salads and can be stuffed or battered and fried. Young leaves and shoots, which have relatively low concentrations of the bitter cucurbitacins, are also important potherbs in Mexican cooking. In India, squash leaves and growing tips are eaten as salad greens or added to vegetable curries. Nineteenth-century Indonesians prepared a dish in which fish and young leaves of *C. moschata* were wrapped in banana leaves and roasted under live coals.

Nutritional Content

Sixty to 85 percent of a mature fresh squash fruit is edible, as compared to over 95 percent edibility in immature fruits. The edible portion of a pepo, which is 85 to 95 percent water by weight, is lacking in most nutrients, particularly protein (0.5 to 2.0 percent) and fat (less than 0.5 percent). Carbohydrates are more concentrated in mature fruits (up to 15 percent of the fresh edible portion) than in the tender fruits of the summer squashes (less than 5 percent). Likewise, calories per 100 grams of edible fresh-weight flesh range from 10 to 25 in summer squashes versus 20 to 45 in the mature fruits known as pumpkins and winter squashes.

The most significant dietary contribution of the pepo is the relatively high concentration of carotenes, the precursors of vitamin A, in cultivars with deep yellow to orange flesh (see Table II.C.8.5).

Table II.C.8.5. Mineral and vitamin content of young fruits (represented by the summer squashes of *Cucurbita pepo*), mature fruits, leaves, and growing tips (including *C. maxima*, *C. moschata*, and *C. pepo*), and ground seed meal (a mixture of *C. pepo* and *C. maxima*); values are per 100 grams of dry seed meal or, in the case of the other structures, of the fresh-weight edible portion

Mineral or vitamin	Immature fruits ^{a,b,c}	Mature fruits ^{a,b,c}	Leaves ^{b,d}	Growing tips ^c	Ground seeds ^e
Potassium (mg)	-	-	-	-	1,111
Magnesium (mg)	-	-	-	-	205
Copper (mg)	-	-	-	-	2.0
Zinc (mg)	-	-	-	-	5.1
Calcium (mg)	14-24	14-50	40,477	-	11.4
Iron (mg)	0.3-0.5	0.4-2.4	0.8, 2.1	31	6.8
Phosphorus (mg)	26-41	21-68	136	-	852
Carotene (mg)	-	0.2-7.8	1.9, 3.6	-	-
Vitamin A (I.U.)	55-450	335-7,810	-	1,000	-
Niacin (mg)	0.4-0.8	0.4-1.0	0.3	1.1	-
Riboflavin (mg)	0.02-0.17	0.01-0.15	0.06	0.21	-
Thiamine (mg)	0.04-0.08	0.05-0.10	-	0.16	-
Ascorbic acid (mg)	5-24	6-45	10, 80	100	-

^aWhitaker and Davis (1962); ^bTindall (1983); ^cMartin (1984); ^dOomen and Grubben (1978); ^eLazos (1986).

Particularly well studied and rich in these and other nutrients are the 'Butternut' and 'Golden Cushaw' cultivars of *C. moschata* and various "hubbard" and "delicious" squashes of *C. maxima*. As a source of vitamin A, these winter squashes compare with sweet potatoes and apricots. Although the raw flesh is higher in vitamins, a half cup of cooked mashed winter squash provides 91 percent of the U.S. Recommended Dietary Allowance (RDA) of vitamin A, 16 percent of the recommended vitamin C, 12 percent of the recommended potassium, 1.7 grams of dietary fiber, low sodium, and only 40 calories. In addition to the carotenoids, squashes are good sources of other compounds with cancer-fighting potential, including flavonoids, monoterpenes, and sterols.

For some nutrients the best source is not the fruit but other parts of the squash plant (see Table II.C.8.5). Leaves are richer in calcium, growing tips provide more iron as well as higher levels of vitamin C and the B vitamins, and seeds contain various minerals including potassium, magnesium, copper, and zinc. Although the nutritional content of flowers has not been studied, the orange petals are undoubtedly rich in carotenes.

Seeds are the most nutritious part of the plant, containing 35 to 55 percent oil and 30 to 35 percent protein by weight. In fact, the naked seeds of 'Lady Godiva' are very similar in agricultural yield and nutritional content to shelled peanuts.

The edible semidrying oil of squash seeds is dark brown with a green tint and a nutty odor. About 80 percent of the oil consists of unsaturated linoleic (40 to 50 percent) and oleic (30 to 40 percent) acids. The dominant saturated fatty acid, palmitic acid, accounts for about 13 percent of oil composition. As with other oilseeds, proteins in squash seeds are rich in nitrogen-containing amino acids such as arginine but lacking in lysine and sulfur-containing amino acids. These proteins are packaged primarily in globulins called cucurbitins. Whereas the testa is highly fibrous, carbohydrates in the decorticated seeds are limited to cell wall cellulose, phytic acid, and a minimal amount of free sugars; starch is absent. Ground seeds (including the testas) are good sources of minerals, particularly potassium, phosphorus, and magnesium (see Table II.C.8.5).

Deena S. Decker-Walters
Terrence W. Walters

Bibliography

- Alefeld, Friedrich. 1866. *Landwirtschaftliche Flora*. Berlin.
- Andres, Thomas C. 1990. Biosystematics, theories on the origin, and breeding potential of *Cucurbita ficifolia*. In *Biology and utilization of the Cucurbitaceae*, ed. David M. Bates, Richard W. Robinson, and Charles Jeffrey, 102-19. Ithaca, N.Y.
- Bailey, Liberty Hyde. 1902. A medley of pumpkins. *Memoirs of the Horticultural Society of New York* 1: 117-24.
1929. The domesticated cucurbits. *Gentes Herbarum* 2: 62-115.
1937. *The garden of gourds*. New York.
1943. Species of *Cucurbita*. *Gentes Herbarum* 6: 266-322.
1948. Jottings in the cucurbits. *Gentes Herbarum* 7: 448-77.
- Bates, David M., Richard W. Robinson, and Charles Jeffrey, eds. 1990. *Biology and utilization of the Cucurbitaceae*. Ithaca, N.Y.
- Bukasov, S. M. 1930. The cultivated plants of Mexico, Guatemala, and Colombia. *Trudy po prikladnoj botanike, genetike i selekcii* 47: 1-533.
- Castetter, E. F. 1925. Horticultural groups of cucurbits. *Proceedings of the American Society of Horticultural Science* 22: 338-40.
- Castetter, E. F., and A. T. Erwin. 1927. A systematic study of squashes and pumpkins. *Bulletin of the Iowa Agricultural Experiment Station* 244: 107-35.
- Conrad, N., D. L. Asch, N. B. Asch, et al. 1984. Accelerator radiocarbon dating of evidence for prehistoric horticulture in Illinois. *Nature* 308: 443-6.
- Cutler, Hugh C., and Thomas W. Whitaker. 1956. *Cucurbita mixta* Pang. Its classification and relationships. *Bulletin of the Torrey Botanical Club* 83: 253-60.
1961. History and distribution of the cultivated cucurbits in the Americas. *American Antiquity* 26: 469-85.
- Decker, Deena S. 1985. Numerical analysis of allozyme variation in *Cucurbita pepo*. *Economic Botany* 39: 300-9.
1988. Origins(s), evolution, and systematics of *Cucurbita pepo* (Cucurbitaceae). *Economic Botany* 42: 4-15.
- Decker, Deena S., and Lee A. Newsom. 1988. Numerical analysis of archaeological *Cucurbita pepo* seeds from Hontoon Island, Florida. *Journal of Ethnobiology* 8: 35-44.
- Decker, Deena S., and Hugh D. Wilson. 1987. Allozyme variation in the *Cucurbita pepo* complex: *C. pepo* var. *ovifera* vs. *C. texana*. *Systematic Botany* 12: 263-73.
- Eisendrath, Erna Rice. 1962. Portraits of plants. A study of the "icones." *Annals of the Missouri Botanical Garden* 48: 291-327.
- Filov, A. I. 1966. Ekologija i klassifikacija tykuy. *Bjulleten' Glavnogo botaniceskogo sada* 63: 33-41.
- Grebensčikov, Igor. 1955. Notulae cucurbitologicae II. *Kulturpflanze* 3: 50-9.
1958. Notulae cucurbitologicae III. *Kulturpflanze* 6: 38-59.
1969. Notulae cucurbitologicae VII. *Kulturpflanze* 17: 109-20.
- Heiser, Charles B. 1979. *The gourd book*. Norman, Okla.
- Kay, Marvin, Francis B. King, and Christine K. Robinson. 1980. Cucurbits from Phillips Spring: New evidence and interpretations. *American Antiquity* 45: 806-22.
- Lazos, Evangelos S. 1986. Nutritional, fatty acid, and oil characteristics of pumpkin and melon seeds. *Journal of Food Science* 51: 1382-3.
- Martin, Franklin W. 1984. *Handbook of tropical food crops*. Boca Raton, Fla.
- Merrick, Laura C. 1990. Systematics and evolution of a domesticated squash, *Cucurbita argyrosperma*, and its wild and weedy relatives. In *Biology and utilization of the Cucurbitaceae*, ed. David M. Bates, Richard W. Robinson, and Charles Jeffrey, 77-95. Ithaca, N.Y.
- Minnis, Paul E. 1992. Earliest plant cultivation in the desert borderlands of North America. In *The origins of agriculture: An international perspective*, ed. C. Wesley Cowan and Patty Jo Watson, 121-41. Washington, D.C.

- Naudin, Charles. 1856. Nouvelles recherches sur les caractères spécifiques et les variétés des plantes du genre *Cucurbita*. *Annales des Sciences Naturelles; Botanique* 6: 5-73.
- Oomen, H. A. P. C., and G. J. H. Grubben. 1978. *Tropical leaf vegetables in human nutrition*. Amsterdam.
- Pangalo, K. I. 1930. A new species of cultivated pumpkin. *Trudy po prikladnoj botanike, genetike i selekcii* 23: 253-65.
- Paris, Harry S. 1989. Historical records, origins, and development of the edible cultivar groups of *Cucurbita pepo* (Cucurbitaceae). *Economic Botany* 43: 423-43.
- Pearsall, Deborah M. 1992. The origins of plant cultivation in South America. In *The origins of agriculture: An international perspective*, ed. C. Wesley Cowan and Patty Jo Watson, 173-205. Washington, D.C.
- Robinson, R. W., H. M. Munger, T. W. Whitaker, and G. W. Bohn. 1976. Genes of the Cucurbitaceae. *HortScience* 11: 554-68.
- Shifriss, Oved. 1955. Genetics and origin of the bicolor gourds. *Journal of Heredity* 36: 47-52.
- Simmons, Alan H. 1986. New evidence for the early use of cultigens in the American Southwest. *American Antiquity* 51: 73-89.
- Sinnott, Edmund W. 1922. Inheritance of fruit shape in *Cucurbita pepo* L. *Botanical Gazette* 74: 95-103.
- Smith, Bruce D. 1992. Prehistoric plant husbandry in eastern North America. In *The origins of agriculture: An international perspective*, ed. C. Wesley Cowan and Patty Jo Watson, 101-19. Washington, D.C.
- Tapley, William T., Walter D. Enzie, and Glen P. Van Eseltine. 1937. *The vegetables of New York*, Vol. 1, Part 4. Albany, N.Y.
- Tindall, H. D. 1983. *Vegetables in the tropics*. Westport, Conn.
- Wall, J. Robert. 1961. Recombination in the genus *Cucurbita*. *Genetics* 46: 1677-85.
- Whitaker, Thomas W. 1931. Sex ratio and sex expression in the cultivated cucurbits. *American Journal of Botany* 18: 359-66.
1947. American origin of the cultivated cucurbits. *Annals of the Missouri Botanical Garden* 34: 101-11.
1951. A species cross in *Cucurbita*. *Journal of Heredity* 42: 65-9.
1956. Origin of the cultivated *Cucurbita*. *American Naturalist* 90: 171-6.
1968. Ecological aspects of the cultivated *Cucurbita*. *HortScience* 3: 9-11.
- Whitaker, Thomas W., and W. P. Bemis. 1964. Evolution in the genus *Cucurbita*. *Evolution* 18: 553-9.
- Whitaker, Thomas W., and Junius B. Bird. 1949. Identification and significance of the cucurbit materials from Huaca Prieta, Peru. *American Museum Novitates* 1426: 1-15.
- Whitaker, Thomas W., and G. W. Bohn. 1950. The taxonomy, genetics, production and uses of the cultivated species of *Cucurbita*. *Economic Botany* 4: 52-81.
- Whitaker, Thomas W., and Hugh C. Cutler. 1965. Cucurbits and cultures in the Americas. *Economic Botany* 19: 344-9.
1971. Prehistoric cucurbits from the valley of Oaxaca. *Economic Botany* 25: 123-7.
- Whitaker, Thomas W., Hugh C. Cutler, and Richard S. MacNeish. 1957. Cucurbit materials from the three caves near Ocampo, Tamaulipas. *American Antiquity* 22: 352-8.
- Whitaker, Thomas W., and Glen N. Davis. 1962. *Cucurbits*. New York.

- Wilson, Gilbert Livingstone. 1917. *Agriculture of the Hidatsa Indians, an Indian interpretation*. Minneapolis, Minn.
- Zhiteneva, N. E. 1930. The world's assortment of pumpkins. *Trudy po prikladnoj botanike, genetike i selekcii* 23: 157-207.

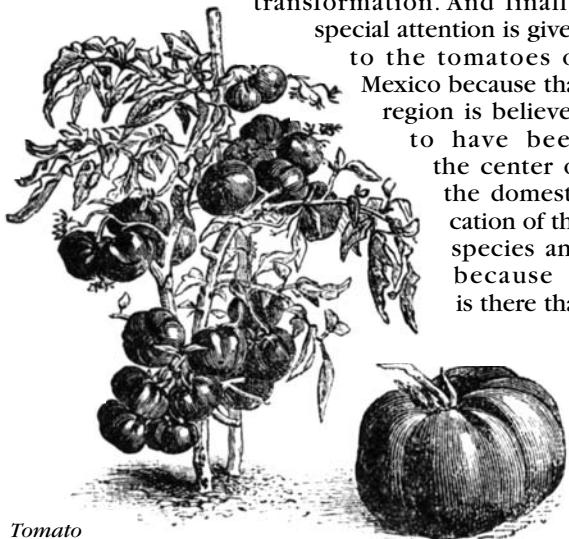
II.C.9 🍅 Tomatoes

The tomato is a perennial plant, generally cultivated as an annual crop. It can be grown in open fields, weather permitting, or in protective structures when temperatures are extreme. In commercial operations, tomatoes are usually planted as a row crop and harvested mechanically when they are still in the green stage. They can also be trained on trellises and harvested throughout most of the year by hand. Tomatoes adapt well and easily to a wide diversity of soils and climates, but they produce best in well-drained soil and temperate climate, with at least a few hours of sunlight each day.

The tomato contains significant amounts of the vitamins A and C, although probably less than the general public has been led to believe. Its importance as a provider of these vitamins depends more on the quantity consumed than on the amount of the vitamins in each fruit. Its vivid color, the fact that it can be used as both a raw and cooked vegetable, and its ability to blend easily with other ingredients has made the tomato a popular international food item and one of the most important vegetables on the world market.

Enormous changes have taken place in the use and distribution of the tomato since the time of its prehistoric origins as a wild, weedy plant. A multidisciplinary research strategy, using archaeological, taxonomical, historical, and linguistic sources is employed in this chapter to trace this remarkable transformation. And finally,

special attention is given to the tomatoes of Mexico because that region is believed to have been the center of the domestication of the species and because it is there that



Tomato

tomatoes have the longest history of use, beginning with the indigenous population.

Taxonomy

The commercial tomato belongs to the genus *Lycopersicon*. It is a relatively small genus within the large and diverse family Solanaceae. The genus is currently thought to consist of the cultivated tomato, *Lycopersicon esculentum*, and seven closely related wild *Lycopersicon* species (Rick 1976: 268; Taylor 1991: 2), all of which are native to northwestern South America. The wild relatives of the cultivated tomato are confined to a narrow coastal area extending from Ecuador to northern Chile and the Galapagos Islands. Some of the wild species contain valuable genes for disease and pest resistance that can be useful for plant breeders in developing new types of cultivated tomatoes when crossed with *L. esculentum*. All of the cultivated tomatoes are derived from the species *L. esculentum*.

The cherry tomato, *L. esculentum* var. *cerasiforme*, is believed to be the direct ancestor of modern cultivated tomatoes and is the only wild tomato found outside South America (Rick 1976: 269). It can also be found in Mexico, Central America, and the subtropics of the Old World (Rick 1976: 269). It bears greater genetic resemblance to the cultivated tomato than other wild species, and the two groups can be freely intercrossed (Taylor 1991: 3).

Lycopersicon esculentum and its close relatives are self-pollinating and exclusively inbreeding due to the position of the stigma inside the anther tube. Wild species may have a slightly exserted stigma, which permits outcrossing, usually with the help of bees or the wind. The modification in the position of the stigma is one of the changes brought about by the domestication process. It is easier to produce a homogeneous product from a self-fertilized plant than one that may cross with a related species.

Although the genus *Lycopersicon* is native to the northwestern coast of South America, there is no archaeological evidence that tomatoes were used by ancient Andean cultures. No plant remains have appeared in site excavations, no clay vessels in the shape of tomatoes have been discovered, and there is no word for the tomato in Quechua or other ancient Andean languages. Such a lack of evidence may indicate that although the tomato existed as a wild species in the Andean region, it was never utilized by pre-Hispanic populations. The commercial tomato in use there at the present time is believed to have been a *post-Columbian* introduction from Mexico after the Americas were unified under Spanish rule. In Mexico the tomato is known by the Spanish name *tomate*, derived from the Nahuatl or Aztec *tomatl*.

As Charles Heiser pointed out some years ago, the theory of the origin of the tomato is strikingly parallel in many ways to that of the chilli pepper, *Capsicum*

spp. (1969: 39). The wild species of both are South American in origin. They reached Mesoamerica¹ at an early date, probably by natural means, and there found a favorable ecological niche, were domesticated, and eventually gave rise, respectively, to the cultivated plants *L. esculentum* and *Capsicum annuum*.

Mexican Tomatoes

The most likely region where the tomato was first domesticated is the Puebla-Veracruz area of Mexico, where according to James Jenkins, the greatest varietal diversity of the cultivated form can be found today. It is thought to have reached this area as a weedy cherry tomato, var. *cerasiforme*, and, upon domestication, to have become the larger-fruited *L. esculentum* (1948: 391, 386). The cherry tomato frequently grows wild as a weed in cultivated fields and is better adapted to wet tropical conditions than any of the other species. It is also used as a cultivated plant and is a popular item in the diet of indigenous peoples. Both wild and cultivated tomatoes have a distinct and independent nomenclature in several Indian languages, indicating an ancient introduction.

We will probably never know how the cherry tomato traveled from the Andean region of the hemisphere to Mesoamerica. Winds or water could have transported the seeds, as could birds who consumed the seeds, then eliminated them at some distant point. Perhaps all of these means of transportation were involved in a kind of stepping-stone journey, with stops along the way where the seeds became plants that reproduced, and new seeds were picked up and moved again and again by such vectors.

Alternatively, humans may have had a hand in the diffusion of the wild ancestor of the tomato. Perhaps it was carried by migrating populations who, in spite of the great distance and geographical barriers between Mexico and South America, were able to move from one area to another. Or again, its introduction to Mexico may have resulted from contact between the two areas that some archaeologists believe was established by seafaring traders as early as 1600 B.C. (Green and Lowe 1967: 56-7).

Still other questions have to do with the extent to which indigenous peoples of Mexico came to accept the tomato and incorporate it into their diets. The plant may have caught the attention of food gatherers because of its general similarity to the green husk tomato, *Physalis* (Jenkins 1948: 392). Unlike the plant we just tracked from South America, this plant is native to central Mexico, where it has a significantly longer tradition of dietary usage than the red tomato. Indeed, there is archaeological evidence of its consumption from 900 B.C. in the excavation in the Tehuacan Valley, Puebla, and from 5090 B.C. in the Valley of Mexico (Smith 1967: 248; Flannery 1985: 266). Basalt grater bowls (*molcayetes*), with incised interiors for grinding vegetal matter, appear in the earliest

stratigraphic levels of the excavation in Tehuacan, and clay bowls began to appear around 1500 B.C. (MacNeish 1967: 290–309). The word *molcajete* comes from the Nahuatl term *molcaxitl*, composed of *molli* (sauce) and *caxitl* (bowl), or “sauce bowl.” One can say with some degree of certainty that they were employed for making salsas of chilli peppers and green (and maybe red) tomatoes, as they are still used in Mexico today.

As in the Andean region, no archaeological evidence of plant remains of the red tomato have been reported from Mesoamerican excavations. In part, this may be because the red tomato is an extremely perishable fruit. However, carbonized seeds are almost indestructible and last indefinitely when they are not mashed or ground up. They can be recuperated from desiccated coprolites (fecal material), which can be reconstituted and returned to their original state in order to be examined. Since coprolites contain the actual materials consumed, analysis of them can provide some important insights into the diet of ancient peoples.

One possible explanation for the absence of tomato seeds in coprolites is that the principal method of preparing tomatoes was by grinding or mashing them in grater bowls or on grinding stones for use in salsas and stews, making the disintegrated seeds impossible to identify in coprolites and other refuse material.

Several changes have taken place in the tomato during the process of domestication. Generally speaking, wild plants have smaller seeds and fruits than the domesticated species. This differentiation can be noted when comparing wild and cultivated tomatoes. Wild tomatoes have two locules, whereas most domesticated fruits are multiloculates, because of an increase in size. Upon domestication, the position of the stigma was established deeper inside the anther tube to insure self-fertilization. Doubt about the extent of such a transformation of the tomato in pre-Columbian times has led some Latin American botanists to view it as a semidomesticated, rather than a fully domesticated plant prior to the arrival of the Europeans. J. León has gone so far as to suggest that it was unimportant as a food crop and considered just another weed in the fields, even though its fruit was the size of some modern varieties (1992: 41–2).

Linguistic evidence is also inconclusive. As previously mentioned, the generic term for the husk tomato in Nahuatl is *tomatl*, with different prefixes or descriptive adjectives used to identify the particular type. The red tomato is known in Mexico by the Spanish term *jitomate* from the Nahuatl *xitomatl*, which may mean “peeled or skinned tomato.” The Nahuatl prefix *xi* is possibly derived from the verb *xipehua*, which denotes “to peel, skin, or flay.” This could be a reference to the calyx that covers the fruit of the husk tomato and is lacking in the red variety. When ancient Mexicans came across the red-fruited tomato

they may have noted its general similarity to the husk tomato and referred to it by a similar name such as *xitomatl*, or “peeled tomato,” to differentiate it from the former fruit.

Unfortunately, sixteenth-century Spanish writers did not distinguish between the *tomatl* and the *xitomatl*; they translated both as *tomate*. Thus, it is impossible to determine which tomato they are referring to unless the Nahuatl text is available. In Bernardino de Sahagún's *The General History of the Things of New Spain*, written in both Nahuatl and Spanish, there are more references to green tomatoes than red, indicating a more frequent use of the former at the time of the European conquest (Sahagún 1951–69).

Nonetheless, all kinds and colors of tomatoes could be bought in the great Tlatelolco market when the Spaniards arrived in Tenochtitlan in 1519. Tomato sellers offered large tomatoes, small tomatoes, green tomatoes, leaf tomatoes, thin tomatoes, sweet tomatoes, large serpent tomatoes, nipple-shaped tomatoes, coyote tomatoes, sand tomatoes, and “those which are yellow, very yellow, quite yellow, red, very red, quite ruddy, bright red, reddish, rosy dawn colored” (Sahagún 1951–69, Book 10: 79). The bad tomato seller was described as one who sold spoiled tomatoes, bruised tomatoes, and those that caused diarrhea (Sahagún 1951–69, Book 10: 68). Clearly, a large variety of tomatoes was for sale in early sixteenth-century Mexico – such a variety that it is impossible to identify some of the types with those on the market today.

Bernal Díaz, who participated in the conquest of Mexico in 1519, related that when the *conquistadors* went through Cholula on their way from Veracruz to Tenochtitlan, the Indians “wanted to kill us and eat our meat” and that “they had their cooking pots ready, prepared with chile peppers, tomatoes and salt . . .” (1980: 148). He also mentioned that the Aztecs ate the arms and legs of their sacrificial victims with a *chimole* sauce, made with chilli peppers, tomatoes, wild onions (*xonacatl*), and salt (1980: 564). The ingredients were nearly the same as that of *salsa mexicana*, in use in most Mexican homes today.

Similarly, stews and salsas, sold on the street or in the markets in sixteenth-century Mexico, were made with red or green tomatoes, chilli peppers, and squash seeds as common ingredients (Sahagún 1951–69, Book 10: 70). Another early visitor to Mexico noted that tomatoes were added to salsas to temper the heat of the chilli peppers (Cervantes de Salazar 1914: 118–11).

The sixteenth-century Jesuit priest José de Acosta, who traveled in Mexico and South America, was no doubt referring to red tomatoes when he described them as fresh and healthy, some being large and juicy, and said they made a tasty sauce and were also good for eating by themselves (1940: 178).

Clearly visitors appreciated the tomato, but it was not until the latter half of the sixteenth century that it became the subject of scientific study. Francisco

Hernandez, the personal physician of Philip II, was commissioned by the king to catalog and describe the medicinal plants being used in New Spain. Hernandez spent the years between 1570 and 1577 traveling throughout the country, preparing a list of the local plants and illustrating them. Unfortunately, his description of the tomato plant gives us little reliable information, because he confused the husk tomato and the red tomato. For example, his chapter on tomatoes is illustrated with a drawing of the former.

Hernandez did note, however, that the tomato was used for medical purposes. The fruit and its juice were used to soothe throat irritations, to treat the discomfort caused by headaches, earaches, and stomachaches, and to ease the pain of mumps (Hernandez 1946, III: 699–715).

There are many references to the production of both species of tomatoes during the colonial period in Mexico. The two were generally planted together, along with chilli peppers, in house gardens, on *chinampas*² in small plots, and in open fields.

Tomatoes were probably geographically limited to Mesoamerica, as no mention of them was made by early Spanish chroniclers who visited the Caribbean. Gonzalo Fernandez de Oviedo, for example, who left the most complete description of New World flora, and whose travels took him to the Caribbean and to parts of South America, but not New Spain, did not mention the tomato.

American Plants Reach Europe

The migration of domesticated plants is closely related to human migration because these plants need human intervention and care to survive. Among other things, many lose their dispersal mechanisms after domestication and cannot be diffused without human help. Unfortunately, plant movements have seldom been considered important enough to warrant registration upon arrival in a new country, which can make the study of plant migration an exercise in frustration for the plant historian.

Many American plants arrived in Iberia in the sixteenth and seventeenth centuries, along with the supposedly more precious cargoes of gold and silver. Some seeds were carried on purpose, perhaps by returning Spaniards, who had become accustomed to the taste of New World foods and flavors; others arrived accidentally, hidden in the nooks and crannies of ships.

Not all of the new plants were well received when they first appeared in Europe. This was especially the case with solanaceous ones such as the tomato, the chilli pepper, and the potato, which were regarded with suspicion and fear by Europeans already familiar with other plants of the same family.

Certainly tomatoes were not an easy ingredient to incorporate, even into the Italian diet where they were later to become a mainstay. They neither looked

nor tasted like any other vegetable known and used by the Italians, and they had a strange texture and consistency. They were too acid to be eaten while green and looked spoiled when they were soft and ripe. They disintegrated upon cooking and were suspected of being poisonous. Thus, it was only after the passage of some considerable length of time that tomatoes were accepted by the Mediterranean peoples to become as much a part of the local food tradition as are wheat, olives, and wine.

But although culinary acceptance of American foods was delayed, European plant specialists, from the very beginning, displayed great interest in any medicinal qualities they might possess. Old diseases, such as plague, were still affecting parts of Europe in the sixteenth century and had been joined by new diseases, like syphilis, to punish populations. Thus, doctors were in constant search of new remedies to treat these ills. They initially had great hopes for the pharmacologic possibilities of New World organisms but soon realized that the new cultivars offered little relief for European illnesses.

However, the American plants did find a place in botanical gardens, popular among scientists of the time, who also established networks through which they exchanged new and exotic plants as well as information about them. In addition, some became popular as ornamentals and could be found in university gardens and on the estates of the nobility. But the tomato had little to offer as an ornamental. Its flowers are a pale yellowish color, not particularly unusual or attractive, and both its leaves and fruit emit a strong, acrid smell that many plant lovers of the time thought offensive.

Herbals with woodcut engravings became popular in the sixteenth and seventeenth centuries, and scientists used them in an effort to establish some order in the plant world. The New World cultivars were quickly fitted in, and much valuable information about them can be gleaned from these publications. In the case of the tomato, the plants appears as a small, heavily ridged and compressed fruit, but large enough to have gone through the domestication process.

American plants in Europe spread out along two different routes after their arrival in Iberia. One led north via Flanders, the other into the Mediterranean via Italy – with all of the former and significant portions of the latter then under Spanish domination, which facilitated communication between these areas.

The moderate climate and loose soil of the Mediterranean countries proved ideal for the adaptation of the tomato as well as other New World plants. They arrived, not as competition for the local cultivars already in production, but as complementary crops whose planting and harvesting schedules did not coincide nor interfere with those of the traditional Mediterranean crops.

Tomatoes in Spain

Spain was doubtless the first stop for the tomato on its migration throughout Europe because Castile held a monopoly on the transport of its New World products to the Continent. Unfortunately, although officials of the Casa de la Contratación³ kept a watchful eye on all cargo unloaded in Seville so as to ensure the collection of royal import taxes, they seldom recorded the arrival of new plants. Thus, there is no record of the arrival of the tomato in Seville – the only port for Spanish ships returning from the New World.

In fact, there are few historical references to the use of tomatoes in sixteenth-century Spain. They may have been adopted first by rural people, who ate them fresh with a little salt like their eighteenth-century peasant descendants in southern Spain (McCue 1952: 327). No Spanish cookbooks were published at this time, however, and there is no mention of tomatoes having been a part of the diet in the early years.

Nor were tomatoes included in sixteenth-century Spanish herbals, although several described and illustrated New World plants. The husk tomato, for example, arrived in the sixteenth century and is known to have been cultivated in botanical gardens in southern Spain. It was listed as an exchange plant from the garden of Dr. Juan Castañeda, a physician at the Flamenco Hospital in Seville, to the Belgian botanist Clusius (Charles de l'Ecluse) at the end of the century (Alvarez Lopez 1945: 276). Castañeda appears to have been Clusius's principal supplier of American and Iberian plants. Clusius made several trips to Spain to obtain information and specimens of new plants, but he did not mention having encountered the tomato on his travels.

The first written reference to the cultivation of the tomato in Spain was penned around the turn of the seventeenth century. It appeared in a small book by Gregorio de Rios, a priest who worked in the botanical garden of Aranjuez, which was supported by the King, Philip II. This book, *Agricultura de jardines, que trata de la manera que se han de criar, gobernar y conservar las plantas*, mentions several American plants. Rudolf Grewe has translated his comments on the tomato as follows: "Tomatoes [*pomates* in the original]: There are two or three kinds. It is a plant that bears some segmented fruits [*pomas aquarteronadas*] that turn red and do not smell. It is said that they are good for sauces. They have seeds, last for two or three years, and require a lot of water. There is a kind said to be from Cairo" (Grewe 1988: 73). By this time at least some Spaniards had apparently adopted the Aztec method of preparing tomatoes in a sauce.

Tomatoes appeared on the list of purchases of the Hospital de la Sangre in Seville in the early seventeenth century. Four pounds were purchased on July 20, and another 2 pounds on August 17, 1608 (Hamilton 1976: 859). The same account lists the purchase

of cucumbers "for making salads," and it is possible that tomatoes were used for the same purpose. This appears to have been the only attempt of the hospital to introduce the tomato into its diet as there seem to have been no further purchases.

In addition to such rather scanty seventeenth-century historical information on the tomato there can be added information from indirect sources. Sixteenth- and seventeenth-century Spanish writers had a fascination with all things from the New World and delighted in borrowing vocabulary from Hispanic Indian languages in their works. Among the words most frequently used were those of fruits and vegetables. Spanish variations of Nahuatl, Quechua, and Caribbean plant names appear in the works of Lope de Vega, Tirso de Molino, Miguel de Cervantes y Saavedra, and Francisco de Quevedo (Morínigo 1946). The new names, including those for the tomato, were used as metaphors, in analogies, or merely for the exotic sounds of the words in poetry and drama.

Painters also found the new fruits and vegetables colorful subjects for still-life paintings that became popular in the sixteenth and seventeenth centuries. Bartolomé Murillo's "The Kitchen of Angels," painted for the Franciscan Convent in Seville, depicts the preparation of a dish using tomatoes and squash, a combination that was to become typically Mediterranean. The seventeenth century witnessed severe economic problems throughout the Mediterranean. In Spain the expulsion of the Moriscos, who had contributed so much to the agriculture of that country, brought a sharp decline in crop production. The resulting famine was joined by the return of the plague, which added to the general misery by considerably reducing the workforce. All of these factors contributed to a severe scarcity of food, which may have encouraged desperate rural peoples to put aside their fear of being poisoned and experiment with tomatoes in their diets.

One suspects this was the case because in the following century it was noted that tomatoes were a common ingredient in the diet of the rich, who ate them because they liked them, and of the poor, who ate them because they had no choice (McCue 1952: 327). Tomatoes were produced in abundance on truck farms and irrigated fields throughout the country, especially in southern Spain, where they could be harvested year-round. Indeed, farmers were eating tomatoes for breakfast, and a plate of fried red tomatoes and peppers constituted the main meal of the day for many (McCue 1952: 327).

Several American plants were fully adopted into the Mediterranean diet in the eighteenth century, and the more abundant and nutritious diet they allowed has been credited by some with bringing about a mid-century increase in the population.

Under the influence of a new and burgeoning merchant class in the eighteenth century, a greater emphasis was placed on simple, regional food and the

use of local ingredients by everyone, not just the peasants. American vegetables fit well into this new culinary style and were included in the diet, not as new and exotic dishes, but as ingredients that added new flavors to traditional dishes such as thick soups, stews, ragouts, and goulash. In Spain, for example, tomatoes were incorporated into gazpacho (an ancient bread soup, probably dating from Roman times), the rice dish, paella, and *bacalao* (salted codfish). In the process, such dishes acquired new appearances as well as new flavors. The Spanish food historian Nestor Lujan has written that some would like to believe that Spanish and Italian cuisines only began with the introduction of the tomato, because so many dishes cannot be made without it (Lujan 1989: 126).

Clearly, then, tomatoes were well established in Spain by the nineteenth century. In that century, reports from Spain described an abundant production of tomatoes on truck farms and gardens in that country. It was noted that tomatoes were eaten raw with salt, formed the base of sauces, and were cooked in various other ways (McCue 1952: 328).

Tomatoes in Italy

Italy was probably the first country to receive the tomato after Spain, since, as already mentioned, there was a strong Spanish cultural influence apparent during the sixteenth century in those parts of Italy under Spanish domination.

Italy proved to be an ideal country for the adaptation of American plants. The climate and soil were similar to that of central Mexico, and the new plants adjusted easily to the area. Initially, however, tomatoes were grown only in pots and kitchen gardens because they needed a well-managed water supply during the setting of the fruit in summer, usually a dry season in the Mediterranean.

Unlike other parts of Europe, where fresh vegetables were considered food for the poor, Italians had (and have) a singular appreciation for them. This may have been a heritage of the Roman Empire, when men preferred light, easily digestible foods that could be eaten in a supine position. The pressure of growing populations during the second half of the sixteenth century was probably also an incentive to try the new foods.

The tomato in Italy was first mentioned by Petrus Andreas Matthiolus. In the first edition of his herbal *Della historia e materia medicinale*, published in Venice in 1544, it was not referred to by name, but in his 1554 edition he gave it the name of *pomi d'oro*. Unfortunately, Matthiolus mistakenly referred to the tomato as a member of the mandrake family, which focused suspicion upon the plant for centuries to come, as many botanists and writers repeated his description of the plant time and time again. In the 1544 edition he described the unnamed fruit as seg-

mented, green at first, and then acquiring a golden color upon ripening, and he noted that it was eaten like the eggplant, fried in oil with salt and pepper. From his description, it seems apparent that the first tomatoes to reach Italy were yellow, although in the 1554 edition, he added that they also ripened in tones of red (McCue 1952: 292).

At about the same time as the second edition of Matthiolus appeared, a Flemish botanist, Rembert Dodoens, published his herbal, *Cruydt-Boeck*, in Antwerp. He was apparently the first to assign to the tomato the name *poma amoris* or "love apple," which was adopted in translation by the French and English. This name gave the tomato a certain reputation as an aphrodisiac, which probably did nothing to discourage its use. The engraving that accompanied his work shows the tomato as a small, irregular, flat fruit with prominent segments and the name *Gulden-Appel*, translated from the Italian *pomi d'oro* (McCue 1952: 299).

Interestingly, the name *poma peruviana* was given the tomato by Piero Antonio Michel in his herbal, *I cinque libri di piante*, published in 1575 (Jenkins 1948: 382). This must have been nothing more than a remarkable coincidence because he surely could not have been aware that Peru was, in fact, the center of origin of the tomato. Like other European botanists of the time, he was not well informed about New World geography and may actually have considered it as just one general area.

In 1572, another sixteenth-century Italian herbalist, Guilandini de Padua, called the tomato the "tumble from Themistitan." It has been pointed out that this designation probably represents a corrupt spelling of Tenochtitlan, the capital city of Mexico, referred to as Temistitan by Hernando Cortés in two of his letters to the Spanish king, written shortly after the Conquest (Jenkins 1948: 382).

We mentioned that in 1554 Matthiolus observed that tomatoes were fried in oil with salt and pepper, like the eggplant. This may be the first recorded description of Italian tomato sauce. However, its first authentic recipe only appeared in 1692 in one of the early Italian cookbooks, *Lo scalco alla moderna*, written by Antonio Latini that was published in Naples (Grew 1988: 74). Apparently the Spaniards had introduced the Aztec method of preparing the tomato in a sauce into Italy, along with the tomato, because a tomato sauce recipe "in the Spanish style" is included in the book. It called for tomatoes, chilli peppers, onion, salt, oil, and vinegar. However, other recipes for tomato sauce were also published that did not ask for peppers, indicating a separation of these two foods in Europe that were so closely linked in Mesoamerican cooking. The tomato, of course, which combined easily with European ingredients and found multiple uses in the diet, became far more important in Mediterranean cooking than peppers.

The careful hands of Italian gardeners improved

the tomato through selective pressures, turning it into a large, smooth, and thicker-skinned fruit than that which had arrived in the sixteenth century. In addition, they developed a manner of prolonging the use of this perishable vegetable by drying it in the sun, which permitted its reconstitution and use throughout the winter. Much later, tomatoes were canned in southern Italy and became an important item of export. Italian emigrants to the United States and Argentina took their food traditions with them and established a demand for the tomato and tomato sauce in the Americas (Casanova and Bellingeri 1988: 165).

Eastern Mediterranean Tomatoes

The botanist Edgar Anderson has credited the Turks with the diffusion of the tomato into the Levant and the Balkan countries. The Turks probably diffused American plants to eastern Mediterranean countries in the sixteenth century when the Ottoman Empire was dominant in the area. They would have become acquainted with the plants in Italian or Spanish ports and taken them to other countries, much as they did when they took the chilli pepper into Hungary in 1526 (Long-Solis 1988: 62). Peppers and maize also became popular items in the diet of Balkan countries, and Fernand Braudel wrote that it was the Turks who introduced rice, sesame seeds, cotton, and maize into the area in the fifteenth and sixteenth centuries (Braudel 1976, II: 779). In addition, Anderson has noted that there is a wide and apparently coherent area, encompassing the Balkans and Turkey and running along the edge of Iran toward Arabia and Ethiopia, where the tomato has been used for centuries in the everyday diet of common people (Anderson, in McCue 1952: 289–348). The culinary legacy of the Turks is still evident in Mediterranean cuisine from Yugoslavia in the east to Algeria in the west. The popular salads made with tomatoes and peppers, known as *peperonata* in Italy, can be found in the diet of every Mediterranean country, with only slight variations.

Tomatoes in the Far East

Tomatoes became an important part of the Chinese diet only during this century, although they were probably carried to China from the Philippines much earlier. The Spaniards arrived in the Philippines from Mexico in 1564, and after establishing their dominion, introduced many Mesoamerican plants.

From there the tomato reached southern China, perhaps as early as the 1500s, where it was given the name *fan chieh* (barbarian eggplant). The name itself suggests an early introduction. Anderson has pointed out that several crops are known in South China by names that combine the adjective *fan* (southern barbarian) with the name of a long-established Chinese

crop (Anderson 1988: 80). Crops with these names were early introductions; New World crops arriving later are known by the more complimentary adjective *hsi*, meaning Western, or *yang*, meaning ocean (Anderson 1988: 94).

African Tomatoes

Tomatoes are an important food product in Africa today, but the question arises as to how long this has been the case and when they first arrived. The most likely answer is that invaders, explorers, missionaries, and traders all played a role in the tomato's introduction.

The food habits of a region often reflect the influence of such outsiders, and certainly, Portuguese explorers and slave traders would have had an early opportunity to participate in such a cultural transfusion. Probably, Arab traders, active in the ports of Mozambique and Angola, were also instrumental in introducing new crops into Africa.

Another common route for plant diffusion in the early centuries was by way of a well-connected network of monasteries and convents in which seeds and plants were exchanged to help feed the personnel of these institutions. In addition, European botanical gardens had a hand in introducing new plants and crops into English, French, and Dutch colonies of Africa and Asia.

Thus, by at least the seventeenth century, the tomato was in cultivation in North Africa, with an English traveler reporting in 1671 that Spanish *tomates* were grown in the common fields in West Barbary (McCue 1952: 330). Several reports of similar cultivation plots were made in the eighteenth century; by the end of the nineteenth century, tomatoes appear to have been widespread throughout the continent.

Tomatoes in the United States

Despite their being native to the Americas, tomatoes had to be introduced into North America from Europe. Although this introduction occurred in the eighteenth century, tomatoes were slow to gain much of a place in the diet until relatively recently. Today, however, tomatoes rank second only to potatoes as the most important vegetable on the U.S. market. They are also the basic ingredient in that most American of sauces, tomato catsup. In recent years, however, Mexican salsa, composed of tomatoes, chilli peppers, onions, and seasoning has become even more popular on the market than catsup.

The main contribution of the United States to the history of the tomato has been the important role it has played in genetic research programs that have contributed to its improvement. The tomato has many characteristics that make it an ideal subject for plant research. It has an ability to produce and prosper in a

diversity of climates and a short life cycle so that it can produce three generations per year under a well-managed program. Tomatoes produce high seed yields. A self-pollinating mechanism practically eliminates outcrossing, although plants can be crossed under controlled conditions. All of these qualities have enabled rapid progress in the improvement of the tomato in the past decades (Rick 1976: 272).

Genetic resources from wild South American species have helped in the development of cultivars that are tolerant to drought, extreme temperatures, and high-salt content in soils and have increased resistance to the diseases and insects that plague tomatoes.

Other improvements are increased crop yields through larger fruit size and an increase in the number of fruits. Improved fruit quality is evident in the shape, texture, color, and flavor of the product. Post-harvest handling has been improved and storage durability increased. The restricted growth gene has been exploited, making mechanical harvesting easier because of the uniformity of the height of tomato plants. Harvesters have become more elaborate and larger, allowing them to harvest at a faster rate.

In addition, the tomato has been an ideal subject for research in genetic engineering, where the majority of such research is carried out on plants, such as the tomato, that are important as staple foods in basic diets around the world. Resistance to certain diseases that have proved difficult to treat and an improvement in the control of fruit ripening and color are some of the aspects being investigated. Important changes in the quality of tomatoes can be expected through genetic engineering in coming years.

Janet Long

Notes

1. The term "Mesoamerica" refers approximately to the area between the state of Sinaloa in northwestern Mexico and Costa Rica, which at the time of the Spanish conquest contained peoples sharing a number of cultural traits.
2. *Chinampas* are highly productive farm plots surrounded on at least three sides by water.
3. The Casa de Contratación, or House of Trade, was founded by royal order in 1503 and located in Seville. The Casa served as an administrative focal point for commercial traffic involving Spanish colonies in the New World.

Bibliography

- Acosta, J. de. 1940. *Historia natural y moral de las Indias*, ed. Edmundo O'Gorman. Mexico.
- Alatorre, A. 1979. *Los 1001 años de la lengua española*. Mexico.
- Alvarez Lopez, E. 1945. Las plantas de America en la botanica Europea del siglo XVI. *Revista de Indias* 6: 221-88.
- Anderson, E. N. 1988. *The food of China*. New Haven, Conn., and London.
- Braudel, F. 1976. *The Mediterranean and the Mediterranean world in the age of Philip II*, Trans. Siân Reynolds. 2 vols. New York.
- Casanova, R., and M. Bellingeri. 1988. *Alimentos, remedios, vicios y placeres*. Mexico.
- Cervantes de Salazar, F. 1914. *Cronica de la Nueva España*. Madrid.
- Díaz del Castillo, B. 1980. *Historia verdadera de la conquista de la Nueva España*. Mexico.
- Flannery, K. V. 1985. Los orígenes de la agricultura en Mexico: Las teorías y las evidencias. In *Historia de la agricultura: Epoca prehispanica-siglo XVI*, ed. T. R. Rabiela and W. T. Saunders, 237-66. Mexico.
- Green, D. F., and G. W. Lowe. 1967. Altamira and Padre Piedra, early preclassic sites in Chiapas, Mexico. *Papers of the New World Archaeological Foundation*, No. 20, Publication No. 15. Provo, Utah.
- Grewe, R. 1988. The arrival of the tomato in Spain and Italy: Early recipes. *The Journal of Gastronomy* 3: 67-81.
- Hamilton, E. J. 1976. What the New World economy gave the Old. In *First images of America: The impact of the New World on the Old*, ed. F. Chiapelli, 2: 853-84. Los Angeles.
- Heiser, C. B., Jr. 1969. Systematics and the origin of cultivated plants. *Taxon* 18: 36-45.
- Hernandez, F. 1946. *Historia de las plantas de Nueva España*, ed. I. Ochoterena, 3 vols. Mexico.
- Jenkins, J. A. 1948. The origin of the cultivated tomato. *Economic Botany* 2: 379-92.
- León, J. 1992. Plantas domesticadas y cultivos marginados en Mesoamerica. In *Cultivos marginados: Otra perspectiva de 1492*, ed. J. E. Hernández Bermejo and J. León, 37-44. Rome.
- Long-Solis, J. 1988. *Capsicum y cultura: La historia del chilli*. Mexico.
- Lujan, N. 1989. *Historia de la gastronomía*. Spain.
- MacNeish, R. S. 1967. A summary of the subsistence. In *The prehistory of the Tebucan Valley*, ed. Douglas D. Byers, 1: 290-309. Austin, Tex.
- McCue, G. A. 1952. The history of the use of the tomato: An annotated bibliography. In *Annals of the Missouri Botanical Garden* 39: 289-348.
- Matthioli, P. A. 1544. *Di pedacio Dioscoride Anazarbeo libri cinque della historia et materia medicinale tradutti in lingua volgare Italiana*. Venice.
- Morínigo, M. 1946. *América en el teatro de Lope de Vega*. Buenos Aires.
- Rick, C. M. 1976. Tomato (Family Solanaceae). In *Evolution of crop plants*, ed. N. W. Simmonds, 268-72. London.
- Sahagún, B. de. 1951-69. *Florentine Codex, the general history of the things of New Spain*, ed. A. J. O. Anderson and C. Dibble. Santa Fe, N. Mex.
- Smith, C. E., Jr. 1967. Plant remains. In *The prehistory of the Tebucan Valley*, ed. Douglas Byers, 1: 220-55. Austin, Tex.
- Taylor, I. B. 1991. Biosystematics of the tomato. In *The tomato crop: A scientific basis for improvement*, ed. J. G. Atherton and J. Rudich, 1-22. London.



II.D

Staple Nuts

II.D.1 🌿 Chestnuts

In the mountainous areas of the Mediterranean where cereals would not grow well, if at all, the chestnut (*Castanea sativa*) has been a staple food for thousands of years (Jalut 1976). Ancient Greeks and Romans, such as Dioscorides and Galen, wrote of the flatulence produced by a diet that centered too closely on chestnuts and commented on the nuts' medicinal properties, which supposedly protected against such health hazards as poisons, the bite of a mad dog, and dysentery.

Moving forward in time to the sixteenth century, we discover that "an infinity of people live on nothing else but this fruit [the chestnut]" (Estienne and Liébault 1583), and in the nineteenth century an Italian agronomist, describing Tuscany, wrote that "the fruit of the chestnut tree is practically the sole subsistence of our highlanders" (Targioni-Tozzetti 1802, Vol. 3: 154).

A bit later on, Frédéric Le Play (1879, Vol. 1: 310) noted that "chestnuts almost exclusively nourish entire populations for half a year; in the European system they alone are a temporary but complete substitution for cereals." And in the twentieth century, the Italian author of a well-known book of plant-alimentation history mentioned that chestnuts not only were collected to be eaten as nuts

but could also be ground into flour for bread making (Maurizio 1932). He was referring to the "wooden bread" that was consumed daily in Corsica until well into the twentieth century (Bruneton-Governatori 1984). Clearly, then, chestnuts have played an important role in sustaining large numbers of people over the millennia of recorded history (Bourdeau 1894).

The Tree

Geographic location has had much to do historically with those who have found a significant part of their diet at the foot of the chestnut tree. The tree tends to stop bearing fruit north of the fifty-second parallel, and its yield in Eurasia satisfies the growers' wishes only south of a hypothetical line drawn from Brittany to Belgrade and farther east to Trabazon, Turkey – the line ending up somewhere in Iran. In Africa, chestnuts grow only in the Maghreb. In North America, there were many chestnut trees before the first decades of the twentieth century, at which time some three billion were destroyed by a blight. Another species of chestnut exists in China, and Japan is on its way to becoming the world's leading chestnut producer.

Chestnuts grow somewhat haphazardly within these geographic limitations. For example, because they dislike chalky soils, they are rare in Greece, except on some sedimentary or siliceous outcrops, where they can become so abundant that they determine place names, such as "Kastania." In addition, the roots of chestnuts tend to decay in badly drained soils, which helps to explain why the trees thrive on hills and mountainsides. Such exacting requirements also help us pinpoint those regions of Portugal, Spain, France, and Italy where populations were long nourished by chestnuts.

It is true that chestnuts are found beyond the geographic limits just outlined. But these are grown for their wood and not for their fruit (chestnut wood is as strong as oak but significantly lighter) – an entirely different method of cultivation. Fruit-producing chestnut trees must be pruned into low broad shapes,



Chestnut

whereas trees for lumber are encouraged to grow tall. In addition, fruit-producing trees require grafting (such as the marrying of hardy to fruit-bearing species) – an activity deemed vital in historical documents (Serre 1600) because the ungrafted tree produces two or three small chestnuts in one prickly pericarp or husk (called a bur) whose only use is for animal feed. Even in our own times, grafting remains necessary as it is practically the only way to avoid the disease enemies of chestnuts that have so menaced the trees since about 1850.

The Nut

After performing the not-so-easy operations of extracting the chestnut from its bur, hard-peel cover, and adhering tannic skin, one has a nourishing nut that is 40 to 60 percent water, 30 to 50 percent glucids, 1 to 3 percent lipids, and 3 to 7 percent protids. In addition, the nut has significant amounts of trace minerals which vary, depending on the soil; and chestnuts are the only nuts to contain a significant amount of vitamin C.

Dried, the chestnut loses most of its water as its caloric value increases. According to the usual conversion table, 100 grams of fresh chestnuts provide 199 calories; dried, they provide almost twice (371 calories) that amount. (For comparative purposes, 100 grams of potatoes = 86 calories; 100 grams of whole grain wheat bread = 240 calories; 100 grams of walnuts = 660 calories.) (Randoin and de Gallic 1976).

When we pause to consider that our sources place the daily consumption of chestnuts by an individual at between 1 and 2 kilograms, we can quickly understand why the chestnut qualifies as a staple food. And like such staples as wheat or potatoes, chestnuts can be prepared in countless ways. Corsican tradition, for example, calls for 22 different types of dishes made from chestnut flour to be served on a wedding day (Robiquet 1835). When fresh, chestnuts can be eaten raw, boiled, baked, and roasted (roasted chestnuts were sold on the streets of Rome in the sixteenth century and are still sold on the streets of European towns in the wintertime).

Chestnuts also become jam and vanilla-chestnut cream, and they are candied. When dried, they can also be eaten raw, but they are usually ground into flour or made into a porridge, soup, or mash (*polenta* in Italy) and mixed with vegetables, meat, and lard. As flour, chestnuts become bread or pancakes and thickeners for stews. Indeed, speaking of the versatility of chestnuts, they very nearly became the raw material for the production of sugar. Antoine Parmentier (that same great apothecary who granted the potato the dignity of human food) extracted sugar from the nuts and sent a chestnut sugarloaf weighing several pounds to the Academy in Lyon (Parmentier 1780). Research on the possibility of placing chestnuts at the center of the French sugar industry was intensified a

few years later during the Continental blockade of the early nineteenth century. Napoleon's choice, however, was to make sugar from beets.

A Chestnut Civilization

That the geographical areas favorable to chestnut trees and their fruits were precisely the areas in which populations adopted chestnuts as a staple food seems obvious enough. But in order to make full use of the opportunity, populations had to create what might be called a "chestnut civilization," meaning that they had to fashion their lives around the trees, from planting the trees to processing the fruits.

Planting

Chestnut trees seldom grow spontaneously. Moreover, pollination rarely occurs wherever the trees grow in relative isolation from one another, and fructification is poor when the tree is not regularly attended. For all these reasons, it is generally the case that the presence of a chestnut tree is the result of human activity, in contrast to a random act of nature. This is clearly so in the case of plantations, or trees whose alignment marks the borders of fields and pathways. But it is also the case with the countless clusters of two or three trees that cast their shadows upon the small hilly parcels of poor tenants.

It is important to note, however, that people do not plant chestnut trees for themselves. Rather, they do it for generations to come because the trees do not begin to bear fruit until they are at least 15 years old, and their yield is not optimal until they are 50 years old: "Olive tree of your forefather, chestnut tree of your father, only the mulberry tree is yours," as the saying goes in the Cévennes (Bruneton-Governatori 1984: 116).

Cultivation

Most of the operations connected with chestnut cultivation involve looking after the trees. This means clearing the brush beneath them and, when possible, loosening the soil; giving water when really necessary; fertilizing with fallen leaves; repairing enclosures to keep away stray animals whose presence could be catastrophic and whose taste for chestnuts is well known; and above all, trimming branches so that they will bear a maximum amount of fruit. Yet, tree care is hardly an exacting task, requiring only 3 to 8 days a year per hectare of trees (Bruneton-Governatori 1984). The trees, of course, would survive without even this minimal care, important only for improving the yield of nuts, which prompted some critics in the nineteenth century to compare chestnuts to manna falling directly from heaven into the hands of lazy onlookers (Gasparin 1863, Vol. 4: 742).

Yet, when all of the exacting and repetitive tasks involved in growing and preparing chestnuts are contemplated, with an absence of mechanization the

common characteristic, chestnutting suddenly seems like very hard work indeed.

Collecting

Efficient collection required that the area under and around the trees be clean so that few chestnuts would be overlooked. Collecting was a manual job, lasting at least three weeks (chestnuts do not fall all at once), and required the efforts of all members of the family. Perhaps half of the burs – the prickly polycarps – open on the tree or when they hit the soil. The other half had to be shelled, often with the bare and calloused hands of those viewed as tough “chestnutters” by fellow workers. Next the fruits were sorted. The very best nuts were sent to market, about 20 percent were judged “throw-outs” for the pigs, and the rest were set aside for domestic consumption.

Chestnut collection was tedious and hard on the back, requiring about 10 hours of labor for an average collection of between 50 and 150 kg per person. An estimate was made that 110 working days were required (100 women-children/days; 10 men/days) to gather the chestnuts from 2 hectares, which would amount to about 5½ tons of fruit (Hombres-Firmas 1838).

Peeling

Fresh chestnuts constituted the bulk of the diet for those who harvested them until about mid-January – about as long as they could safely be kept. But before they could be eaten, the nuts had to be extracted from their rigid shell and stripped of their bitter and astringent skin. This is a relatively easy procedure when chestnuts are roasted, but generally they were boiled. Peeling chestnuts was usually done by men in front of the fire during the long evenings of autumn and winter. To peel 2 kg of raw chestnuts (the average daily consumption per adult in the first part of the nineteenth century) required about 40 minutes. Therefore, some three hours, or more, of chestnut peeling was required for the average rural family of five. The next morning around 6 A.M. the chestnuts, along with some vegetables, were put into a pot to begin boiling for the day’s main meal.

Drying

The only way to preserve chestnuts for longer periods was to dry them. The method was to spread out the fruit on wattled hurdles high over the heat and smoke of a permanent fire for about two weeks, often in wooden smoking sheds built specifically for this purpose. Following this step, the dried chestnuts – from 5 to 10 kg at a time – were wrapped in a cloth and rhythmically thrashed against a hard surface to separate the nuts from shells and skins that the drying process had loosened.

Dried chestnuts had the effect of liberating peasants from the irksome chore of daily peeling, and the drying procedure had important social consequences

as well. Diego Moreno and S. de Maestri (1975) have noted that the expanding cultivation of chestnut trees in the sixteenth-century Apennines gave birth to hamlets that sprang up around the smoking sheds.

Grinding and Flour

After the chestnuts were dried, they could be ground into flour that would keep for two or three years, provided it was not subjected to moisture. From this flour pancakes and bread were made, although because chestnut flour does not rise, many commentators refused to call the loaves bread. There were also others who had harsh words for other chestnut products, making fun of “this kind of mortar which is called a soup” (Thouin 1841: 173) or that bread which “gives a sallow complexion” (Buc’hoz 1787: 126).

Chestnut Consumers

Chestnuts were mostly the food of rural peasants in mountainous regions that stretched in a belt from Portugal to Turkey. But they were a well-appreciated food by many accounts, such as those of regionalist connoisseurs who praised the “sweet mucilage” (Roques 1837) and the following 1763 text published in *Calendriers . . . du Limousin*:

All the goods nature and art lavish on the table of the rich do not offer him anything which leaves him as content as our villagers, when they find their helping of chestnuts after attending their rustic occupations. As soon as they set eyes on them, joy breaks out in their cottages. Only mindful of the pleasure they then taste, they are forgetful of the fatigues they endured: they are no more envious of those of the towns, of their abundance and sumptuousness (*Calendriers . . . du Limousin* 1763, reprinted in Bruneton-Governatori 1984: 462).

This is not to say, however, that only peasants ate chestnuts, and, in fact, numerous sources indicate that this foodstuff could be a prized dish at higher levels of society. For example, a French nobleman (Michel de Montaigne 1774) recorded that on October 22, 1580, while on his way to Italy, he ordered raw chestnuts. And, in fact, a Spanish nobleman wrote in his account of a campaign against the Moriscos that the whole company, nobility included, consumed 97.4 tons of bread, 33,582 liters of wine, and 240 tons of chestnuts, as against only 19.3 tons of biscuit and 759 kg of chickpeas (Vincent 1975).

We know that chestnuts were served in Utrecht in 1546 at the royal Golden Fleece banquet, and we have the delightful Marie Marquise de Sévigné (1861, Vol. 2: 133–4) playing the woman farmer who claimed to be “beset with three or four baskets” (of chestnuts): “I put them to boil; I roasted them; I put them in my pocket; they appear in dishes; one steps on them.”

This and other quotations tend to obscure the fact that, for the rich in particular, there were chestnuts and then again, there were chestnuts. The French (and the Italians) have two words for chestnut. The ordinary chestnut is called *châtaigne*, whereas the best (and sweetest) chestnut is called a *marron* (which in English is known as the Spanish chestnut). The difference lies in size and form. Usually the husk holds only one *marron* with no dividing skin (the kernel is whole), whereas there may be three or more *châtaignes* in a husk divided by partitions. *Marrons* are the material of commercial candied chestnuts and have historically commanded a price three or four times greater than their common, flawed counterparts. One of the reasons is that the yield of *marrons* is less. Thus, in times past, those who grew them were usually located on a commercial artery and did not depend on chestnuts alone to feed families and pigs.

From the Renaissance on, there were three major commercial roads for chestnuts in Europe. One ran from the Portuguese provinces of Minho and Tras-os-Montes to the harbors of northern Portugal and Galicia where chestnuts were loaded aboard ships, usually bound for Bordeaux. In that port the Iberian chestnuts were combined with chestnuts bought on the Périgueux market and then sent on to Great Britain and the Netherlands. A British author writing of this trade route said that the choicest chestnuts were those grown in Spain or Portugal (Miller 1785).

The French, by contrast, thought the best chestnut was the so-called Lyon chestnut, which was actually an Italian chestnut traveling the second of the three European chestnut arteries. Lyon monopolized the importation of Italian chestnuts, transshipping them to Paris and points farther north. The third route, which also originated in Italy, ran from Milan and Bergamo north to the Germanic countries.

Fresh chestnuts, as we have seen, are perishable, staying fresh for only about three months. And weeks of travel in wagons and the holds of ships did them no good. Thus, transporting chestnuts in bulk was a risky business, and wholesalers fixed their prices accordingly. Only the best chestnuts were shipped, and they went mostly into sweetmeats. In markets they were so costly that only the well-off could purchase them for a tidbit at the table. Consequently, the chestnut trade never did involve large quantities, and most of the chestnuts sold for consumption went through local markets and merchants. In 1872, for example, Paris received barely 6,000 tons of an estimated national crop of 500,000 tons.

The bulk of any chestnut crop, of course, reached no market but was consumed by the peasant families that grew them, along with their poultry and two or three hogs. The British agronomist Arthur Young, who traveled in Limousin, France, during the years 1787–89, calculated that an acre with 70 chestnut trees would feed one man for 420 days or 14 months (Young 1792). This seems a substantial overestimation

of the average number of trees per acre. It was generally the case that between 35 and 100 trees grew on 1 hectare (about 2½ acres). If, however, one assumes that a family living on a hilly and not particularly productive hectare of land could harvest about 2,800 kg of chestnuts, then certainly the chestnuts alone could feed a family for more than half a year. With an average daily consumption of 2 kg per person or 10 kg for a family of five, the 2,800 kg of chestnuts would have fed the family for close to 7 months and a pig or two (350 kg are required to fatten a pig from 100 to 200 kg). The pigs, in turn, might be sold or slaughtered, and one suspects that several pigs on a chestnut farm were a food index of chestnut surpluses.

Chestnuts in Decline

A very good question is why such a useful and valuable foodstuff as chestnuts has today been virtually forgotten. The “golden age” of the chestnut, which seems, in retrospect, to have begun with the Renaissance, had all but vanished by the middle of the nineteenth century (Pitte 1979). It is difficult to quantify the decline because the statistics do not reflect domestic production for self-sufficiency. Nonetheless, a series of events that had a considerable impact on chestnutting can be identified.

One of the first blows dealt to chestnut production (especially in France) was the very hard winter of 1709. According to observers, tree loss was considerable, even to the point of discouraging replanting (*Journal Économique* 1758). The *Intendant* in Limoges reported in 1738 that owners there had not replanted even a twentieth of the trees that had frozen 29 years earlier. And in 1758, a chestnut plantation around the Pau castle was uprooted. Unquestionably, the winter of 1709 caused considerable concern for the future of chestnut cultivation, as did the similarly devastating winters in 1789 and 1870.

A second factor was the substitution of mulberry trees for chestnuts around the Rhone valley, where Lyon and its silk industry exerted considerable influence. Silkworms are fond of mulberry leaves, and the mulberry tree (unlike the chestnut) grows fast and produces quickly. Its cultivation, therefore, encouraged a cash economy as opposed to self-sufficiency.

A third reason for the decline of the chestnut, at least in France, may have been free trade in wheat. In 1664, fear of food shortages had prompted Colbert to take the severe measures of controlling wheat production and prohibiting its exportation. At the same time, the exportation of chestnuts was encouraged. Such regulations lasted about a century before the free traders triumphed over regional monopolists and wheat became a cheap and widely available foodstuff, even competing with chestnuts in regions that had traditionally grown them.

Chestnuts also came under fire beginning in the eighteenth century as a foodstuff deficient in nutri-

ents. A well-off society that tasted a *marron* occasionally pitied the unfortunate peasants who were condemned to gulping down a pigfood – the *châtaigne*. Such a diet represented “The International of Misery and Chestnut,” according to Leroy Ladurie (1966).

But this was the time of the Physiocrats, who thought the soil was the only source of wealth and aimed at improving the productivity of farming by questioning all traditional rural economic processes. That chestnuts suffered at their hands is undisputable. In a query sent to provincial learned societies, François Quesnay and Victor Riqueti Mirabeau, both initiators of the Physiocratic school, asked the following questions: “Are there acorns or chestnuts used as foodstuff for pigs? Do chestnuts give a good income? Or are said chestnuts used as food for the peasants, inducing them to laziness?” (Quesnay 1888: 276). And in an agricultural text of a few decades later, the question of laziness was pursued: “To my knowledge, inhabitants of chestnut countries are nowhere friendly with work” (Bosc and Baudrillard 1821: 272). It went on to suggest that they refused to replace their trees with more productive plants because of their fear of taxation and concluded that they were not worthy citizens of the modern state.

Interestingly, the voice of François Arouet Voltaire (1785: 106) was one of the few who defended the chestnut:

[W]heat surely does not nourish the greatest part of the world. . . . There are in our country, whole provinces where peasants eat chestnut bread only; this bread is more nourishing and tastier than the barley or rye bread which feeds so many people and is much better for sure than the bread ration given to soldiers.

More than two hundred years later we find A. Bruneton-Governatori (1984) agreeing with Voltaire, noting that chestnuts provide a balanced diet and around 4,000 calories of energy. The condemnation the chestnut received in the eighteenth and nineteenth centuries might “raise doubts about the pertinence of contemporary evidence concerning the nutrition of non-elite people.”

The half century from 1800 to 1850 was one of slow decline for the European chestnut as fewer and fewer people were interested in cultivating it, eating it, or defending it. One notes 43,000 trees uprooted in the Italian Piedmont between 1823 and 1832, and public surveyors here and there reported that chestnut-planted lands were diminishing. But following the midpoint of the nineteenth century, we have statistics in France that demonstrate vividly the magnitude of the decline. In 1852, there were 578,224 hectares of land given to chestnut cultivation; in 1892, 309,412; in 1929, 167,940; and in 1975, only 32,000 (Bruneton-Governatori 1984).

A final factor in the decline of chestnuts was doubtless the so-called ink disease, which officially

began in Italy in 1842, had spread to Portugal by 1853, and reached France by 1860. The disease could kill chestnut trees in two or three years, and entire hectares of dried-up trees discouraged any notions of replanting. And, as mentioned, another disease appeared in North America to kill practically all the chestnuts there.

Thus, chestnuts went the way of so many other foods of the past as, for example, salted codfish. Once popular and cheap foods that fed many, they have now become expensive delicacies for a few.

Antoinette Fauve-Chamoux

Bibliography

- Arbuthnot, John. 1732. *Practical rules of diet in the various constitutions and diseases of human bodies*. London.
- Bolens, Lucie. 1974. *Les méthodes culturelles du Moyen Âge d'après les traités d'économie andalous*. Geneva.
- Bosc, Louis, and Jacques Baudrillard. 1821. *Dictionnaire de la culture des arbres et de l'aménagement des forêts*. In Abbé Henri Tessier and André Thouin. *Encyclopédie méthodique, agriculture*, t.VII. Paris.
- Bourdeau, Louis. 1894. *Histoire de l'alimentation, substances alimentaires, procédés de conservation, histoire de la cuisine. Études d'histoire générale*. Paris.
- Bruneton-Governatori, Ariane. 1984. *Le pain de bois. Ethno-histoire de la châtaigne et du châtaignier*. Toulouse.
- Buc'hoz, Pierre-Joseph. 1770. *Dictionnaire universel des plantes, arbres et arbustes de la France*. Paris.
1787. *L'art de préparer les aliments suivant les différents peuples de la terre*. Paris.
- Calendriers Éclésiastiques et civils du Limousin. 1763. *Observations sur le châtaigner et les châtaignes*. Limoges.
- Estienne, Charles, and Jean Liébault. 1583. *L'agriculture et maison rustique*. Paris.
- Gasparin, Comte Adrien de. 1863. *Cours d'agriculture*. 6 vols. Paris.
- Hombres-Firmas, Baron Louis d'. 1838. *Mémoire sur le châtaignier et sa culture dans les Cévennes (1819)*. Published in *Recueil de mémoires et d'observations de physique, de météorologie, d'agriculture et d'histoire naturelle*. Nîmes, France.
- Jalut, Guy. 1976. Les débuts de l'agriculture en France: Les défrichements. *La préhistoire française*, Vol. 2: 180–5. Paris.
- Journal économique ou mémoires, notes et avis sur les arts, l'agriculture et le commerce*. . . . 1758. Paris.
- Le Play, Frédéric. 1879. *Les ouvriers européens*. 6 vols. Paris.
- Le Roy Ladurie, Emmanuel. 1966. *Les paysans du Languedoc*. Paris.
- Maurizio, Adam. 1932. *Histoire de l'alimentation végétale depuis la préhistoire jusqu'à nos jours*. Paris.
- Miller, Philip. 1785. *Dictionnaire des jardiniers*. Paris.
- Montaigne, Michel de. 1774. *Journal de voyage 1580–1581*. Rome and Paris.
- Moreno, D., and S. de Maestri. 1975. Casa rurale e cultura materiale nelle colonizzazione dell'Appennino genovese tra XVI e XVII secolo. *I paesaggi rurali europei. Deputazione di storia patria per l'Umbria*, Bolletino N° 12, Perugia.
- Parmentier, Antoine. 1780. *Traité de la châtaigne*. Bastia, Corsica.

- Pitte, Jean-Robert. 1979. L'homme et le châtaignier en Europe. In *Paysages ruraux européens*. Travaux de la conférence européenne permanente pour l'étude du paysage rural, Rennes-Quimper, 26–30 Sept. 1977.
- Quesnay, François. 1888. *Oeuvres économiques et philosophiques*, ed. A. Oncken. Frankfurt and Paris.
- Randoin, Lucie, and Pierre de Gallic. 1976. *Tables de composition des aliments*. Paris.
- Robiquet, François-Guillaume. 1835. *Recherches historiques et statistiques sur la Corse*. Paris.
- Roques, Joseph. 1837. *Nouveau traité des plantes usuelles spécialement appliqué à la médecine domestique et au régime alimentaire de l'homme sain ou malade*. Paris.
- Serre, Olivier de. 1600. *Le théâtre d'agriculture et mesnage des champs*. Paris.
- Sévigné, Marie Marquise de. 1861. *Lettres*. 11 vols. Paris.
- Targioni-Tozzetti, Ottaviano. 1802. *Lezioni di agricoltura, specialmente toscana*. 4 vols. Florence, Italy.
- Thouin, André. 1841. *Voyage dans la Belgique, la Hollande et l'Italie (1796–1798)*. Paris.
- Vincent, Bernard. 1975. Consommation alimentaire en Andalousie orientale (les achats de l'hôpital royal de Guadix). *Annales E.S.C.* 2–3: 445–53.
- Voltaire, François Arouet dit. 1785. *Dictionnaire philosophique*, Vol. 48 in *Oeuvres complètes*. 92 vols. Paris.
- Young, Arthur. 1792. *Travels during the years 1787, 1788 and 1789*. London.

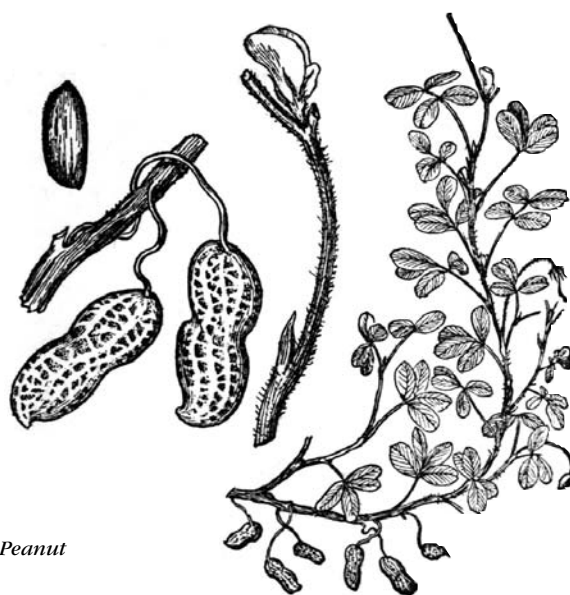
II.D.2 🐾 Peanuts

Peanut or groundnut (*Arachis hypogaea* L.) is a major world crop and member of the Leguminosae family, subfamily Papilionoidae. *Arachis* is Greek for “legume,” and *hypogaea* means “below ground.” *Arachis*, as a genus of wild plants, is South American in origin, and the domesticated *Arachis hypogaea* was diffused from there to other parts of the world. The origin of *Arachis hypogaea* var. *hypogaea* was in Bolivia, possibly as an evolutionary adaptation to drought (Krapovickas 1969). Certainly the archaeological evidence of the South American origins is secure. However, the debate about the pre-Columbian presence of New World plants in Asia (especially India) remains unresolved. The other species of *Arachis* that was domesticated prehistorically by South American Indians was *A. villosulicarpa*, yet the latter has never been cultivated widely.

As the peanut's nutritional and economic importance became recognized, it was widely cultivated in India, China, the United States, Africa, and Europe. Thus, the peanut is another of the New World food crops that are now consumed worldwide. The peanut is popular as a food in Africa and in North America, especially in the United States; peanut-fed pigs produce the famous Smithfield ham of Virginia, and peanut butter is extremely popular. There is also much interest in peanut cultivation in the United States.

Botanically, the varieties of peanuts are distinguished by branching order, growth patterns, and number of seeds per pod. The two main types of peanuts, in terms of plant growth, are “bunch or erect,” which grow upright, and “runners or prostrate,” which spread out on or near the ground. Commercially, peanuts are grouped into four market varieties: Virginia, Runner, Spanish, and Valencia. The former two include both bunch and runner plants, and the latter two are bunch plants. Details on the life cycle and growth of the peanut and its harvesting are provided later in this chapter (Lapidis 1977). Table II.D.2.1 shows the various characteristics of the four varieties.

Peanuts are also called “groundnuts” because they are not true tree nuts. Peanuts are seeds of tropical legumes with pods that grow underground to protect the plant's seeds from seasonal drought and from being eaten by animals. Peanuts consumed by humans are dried seeds of the Leguminosae family, as are kidney, pinto, lima, and soy beans, as well as peas and lentils. The dried shell of the peanut corresponds to bean and pea pods. The names “peanut” and “ground pea” (as the food was called when it was first eaten in North America) became popular because the dried seed had a nutlike shell and texture, and it looked like a pea. Peanuts are also named “goobers,” “earth almonds,” “earth nuts,” “Manila nuts,” “monkey nuts,” “pinda,” and *pistache de terre*. But these terms sometimes also apply to other plants of similar character, such as *Voandzeia subterranea*, found in West Africa, Madagascar, and South America, and the “hog peanut” (*Amphicarpaea menziesii*), found in North America.



Peanut

Table II.D.2.1 *Characteristics of peanut varieties*

Variety	Origin	Season (days)	Nut
Virginia (<i>Arachis hypogaea</i> var. <i>hypogaea</i>)	Probably originated in Amazon	140–160	Large; long; slender; 1–2/pod
Runner	A cross between Virginia and Spanish	140–160	Small to large; stubby; 2–3/pod
Spanish (<i>Arachis hypogaea</i> var. <i>vulgaris</i>)	Originated in Brazil	90–120	Small; round; 2–3/pod
Valencia (<i>Arachis hypogaea</i> var. <i>fastigiata</i>)	Originated in Brazil and Paraguay	90–120	Small; 3–6/pod

Source: Adapted from Ockerman (1991), p. 546.

Structure

The peanut plant roots at its nodes and is self-pollinating, with flowers that open and die after fertilization. It is unique in that it flowers above the ground, but after fertilization the fruit develops below ground in the soil.

The peanut consists of the germ (heart), two cotyledons (halves of the peanut), the skin, and the shell. The pod is tough and stays closed as the seeds ripen, but the seeds themselves have soft, digestible coats. They have been eaten by humans ever since South American Indians first domesticated them during prehistoric times.

Origins and History

Evidence exists of peanuts having been grown in Peru as early as 2000 B.C. (Sauer 1993). As mentioned, they are believed to have originated in South America, and many wild species of the genus *Arachis* are found there. Spanish explorers spread the peanut to Europe and the Philippines, and Portuguese explorers took it to East Africa. It reached North America circuitously via the slave trade from Africa, although in pre-Columbian times it probably came to Mexico from South or Central America. The stocks developed in Africa provided the basis for many varieties now grown in the United States. Initially, peanuts were cultivated in the United States for livestock feed to fatten farm animals, especially pigs, turkeys, and chickens. But they gained commercial importance after the Civil War, with much of the credit due to George Washington Carver at the Tuskegee Institute. One of America's most distinguished African Americans of the nineteenth century, Carver spent his life developing various uses for peanut products, and they became important as a food and as an oil source.

In addition, commercial mills that crushed peanuts for oil were developed independently in Asia and Europe. Europe's inability to meet a demand for olive oil led to a market for peanut oil there, with the peanuts coming mainly from West Africa, and then from India after the opening of the Suez canal. Peanuts subsequently were cultivated in all tropical and subtropical parts of the world.

Unique Characteristics

Unlike most legumes, peanuts store oil instead of starch. During the early growth of the cotyledon storage cells (up to about 30 days after the peg or gynophore strikes the soil), starch granules predominate, and lipid and protein bodies are few. After this stage, however, to about 45 days, both lipid and protein bodies increase rapidly, and from 45 to 68 days, protein bodies and especially lipid bodies continue to expand. The plant's protein and fat come from these bodies in the peanut cotyledon. In the final stage, there is little further growth of the cotyledon (Short 1990; Weijian, Shiyao, and Mushon 1991). Most peanuts require 140 to 150 frost-free days to mature, but such factors as growing season, location, and time of fruit set also influence the time required to reach maturity (Cole and Dorner 1992). Table II.D.2.1 gives characteristics of peanut varieties.

Peanut Pathogens and Pests

Approximately a quarter of the peanut fruit and vine crop is lost because of plant disorders wrought by insects, bacteria, fungi, nematodes, and viruses. *Sclerotinia minor*, the cause of sclerotinia blight, and *Cercospora arachidicola*, the cause of early leaf spot, are two important peanut pathogens. These are controlled by herbicides. Unfortunately, resistance to one is often associated with high susceptibility to the other, and resistance to *S. minor* is also associated with small seed size and an undesirable shade of tan color for the Virginia peanut type (Porter et al. 1992). Bacterial wilt is caused by *Pseudomonas solanacearum*. Fungal species, including *Aspergillus*, *Rhizopus*, *Fusarium*, and others, cause various diseases.

The peanut root-knot nematode (*Meloidogyne arenaria* [Neal] Chitwood race 1) is a major pest in the peanut-producing areas in the southern United States. These microscopic worms greatly reduce yields but can be controlled with fumigants and nematicides. Efforts are now moving forward to select *M. arenaria*-resistant species of peanuts, because chemical controls of the pest are becoming more limited (Holbrook and Noe 1992).

Tomato spotted wilt virus (TSWV) decreases seed

number and weight. Other viruses cause such diseases as spotted wilt and chlorotic rosettes. Insects that attack peanuts include the corn rootworm, which causes rot, and the potato leafhopper, which secretes a toxic substance, damaging the leaves.

Staphylococcus aureus brings about microbial degradation of fat in peanuts, but the major pathogen with relevance to human health is a fungal aflatoxin. It is a carcinogenic metabolite of *Aspergillus flavus* and *Aspergillus parasiticus*, which may cause or promote liver cancer in humans, especially when infected nuts are eaten in large quantities.

Although neither pathogen nor pest, drought is another major limiting factor in peanut production, and efforts are now progressing to develop drought resistance in some varieties (Branch and Kvien 1992).

Horticulture

Production

The six leading peanut-producing countries of the world are India, China, the United States, Nigeria, Indonesia, and Senegal. World production of peanuts in the shell for 1992 was 23 million metric tons, with Asia and Africa producing 90 percent of the total (FAO 1993).

In the United States, the state of Georgia leads the nation in peanut production, followed by Texas, Alabama, and North Carolina (United States Department of Agriculture 1992). The most famous peanut producer in the United States is former President Jimmy Carter.

Cultivation

Peanuts need hot climates with alternating wet and dry seasons and sandy soils. Ideally, rainfall or moisture from irrigation should total at least an inch a week during the wet season. Peanuts are planted after the danger of frost is gone, when soil temperatures are above 65° F. The soil is usually treated with herbicides, limed, fertilized, and plowed before planting. Insecticides may then be applied, and herbicides are applied between preemergence and cracking time (postemergence). Postemergence practices involve cultivation, insecticides if needed, and herbicides for weed control. Calcium sulfate is provided to maximize peanut fruit development. This addition of calcium is important in peanut fertilization, because insufficient levels can lead to empty pods with aborted or shriveled fruit (Cole and Dörner 1992). Peanuts are usually rotated with grass crops such as corn, or with small grains, every three years. This rotation reduces disease and soil depletion. Efforts have been made to intercrop peanuts with other plants, such as the pigeon pea or cotton, but these have not been successful.

Harvesting

Only about 15 percent of peanut flowers produce fruit. The harvest includes both mature and immature varieties, as all fruits do not mature at the same time,

and about 30 percent is immature at harvesting. Maturity can be estimated in a variety of ways. The "shell-out method" for recognition of maturity has to do with the darkening of the skin (testa) and the inside of the hull. The "hull scrape method" is done by scraping the outer shell layer (exocarp) to reveal the color of the middle shell (mesocarp), which is black in the mature peanut. Peanut harvesting involves removing plants from the soil with the peanuts attached (the upright plant is better suited to mechanical harvesting). A peanut combine is used to remove the pods from the plant.

Storage

After harvesting, the peanuts are cleaned by removing stones, sticks, and other foreign material with a series of screens and blowers. For safe storage the peanuts are dried with forced, heated air to 10 percent moisture.

Cleaned, unshelled peanuts can be stored in silos for up to six months. Shelled peanuts are stored for a lesser time in refrigerated warehouses at 32–36° F and 60 percent relative humidity, which protects against insects. A high fat content makes peanuts susceptible to rancidity, and because fat oxidation is encouraged by light, heat, and metal ions, the fruit is best stored in cool, dry places (McGee 1988). On the whole, however, unshelled peanuts keep better than shelled.

Processing

Peanuts may be processed shelled or unshelled, depending on the desired end product. Those left unshelled are mainly of the Virginia and the Valencia types. They are separated according to pod size by screening; discolored or defective seeds are removed by electronic color sorting, and the stems and immature pods are removed by specific gravity (Cole and Dörner 1992). Peanuts that are to be salted and roasted in the shell are soaked in a brine solution under pressure, and then dried and roasted.

Peanuts to be shelled are passed between a series of rollers, after which the broken shells and any foreign materials are removed by screens and blowers. Next, the shelled peanuts are sorted by size. Any remaining foreign materials and defective or moldy seeds are removed by an electronic eye, which inspects individual seeds. Ultraviolet light is useful for detecting aflatoxin contamination.

Peanuts are frequently blanched to remove the skins and hearts. This can be done by roasting (259–293° F for 5 to 20 minutes), or by boiling, after which they are rubbed to remove the skins. Then the kernels are dried to 7 percent moisture and stored, or converted into various peanut products.

Another method – dry roasting – is popular because it develops a desirable color, texture, and flavor for peanut butter, candies, and bakery products. In

this case, unblanched peanuts are heated to 399° F for 20 to 30 minutes, then cooled and blanched.

Shelled peanuts are usually dry roasted in a gas-fired rotary roaster at 399° F, then cooled to 86° F, after which they are cleaned and the skins removed for making peanut butter. Oil-roasted peanuts are placed in coconut oil or partially hydrogenated vegetable oil at 300° F for 15 to 18 minutes until the desired color is achieved, whereupon a fine salt is added. Roasting makes the tissue more crisp by drying and also enhances flavor because of the browning reaction. Relatively low temperatures are used to avoid scorching the outside before the inside is cooked through. The roasting of peanuts also reduces aflatoxin content. For example, roasting for a half hour at 302° F may reduce aflatoxin B₁ content by as much as 80 percent (Scott 1969).

And finally, peanut oil is extracted by one of three different methods: hydraulic pressing, expeller pressing, or solvent extraction.

Food Uses

Traditionally, peanuts were used as a source of oil and, even now, most of the world's peanut production goes into cooking oils, margarines, and shortenings, as well as into the manufacture of soap and other industrial processes. Also called "arachis oil," "nut oil," or "groundnut oil," peanut oil is a colorless, brilliant oil, high in monounsaturates. Virgin oil is mechanically extracted (expeller pressed at low temperature [80–160° F]), and lightly filtered. This method provides the lowest yield but the highest-quality edible oil.

Refined oil is typically produced by solvent extraction. It is made from crushed and cooked peanut pulp, which is then chemically treated in order to deodorize, bleach, and neutralize the flavor of the oil. In the United States, only low-grade nuts are used for oil production. The fatty acid composition is quite variable for a number of reasons, such as genotype, geography, and seasonal weather (Holaday and Pearson 1974). When refined oil is stored at low temperature, a deposit is formed, and hence it cannot be used in salad oils and dressings.

Only peanuts that are free from visible mold and subject to less than 2 percent damage are used for edible purposes. In the United States and Western Europe, most peanuts to be eaten go into the "cleaned and shelled" trade and are consumed as roasted and/or salted nuts, as peanut butter, or as a component of confections.

Because of its high protein and low carbohydrate content, peanut butter was first developed in 1890 as a health food for people who were ill. It is a soft paste made from Virginia, Spanish, or other types of peanuts. The skin and germ are removed, and the kernels are dry roasted and ground. Salt, antioxidants, flavors, and sugars (dextrose or corn syrup) may be added after

grinding. Hydrogenation and/or the addition of emulsifiers prevents separation. "Crunchy style" peanut butter has bits of roasted nuts mixed into it.

Peanut butter is approximately 27 percent protein, 49 percent fat, 17 percent carbohydrate, 2 percent fiber, and 4 percent ash. Its sodium content is approximately 500 mg per 100 g. Peanut butter has good stability even after two years of light-free storage at 80 degrees Fahrenheit (Willich, Morris, and Freeman 1954), and keeps longer if refrigerated. But sooner or later, it becomes stale and rancid.

Peanuts are frequently employed in the cuisines of China, Southeast Asia, and Africa. The residual high-protein cake from oil extraction is used as an ingredient in cooked foods and, in Chinese cooking, is also fermented by microbes.

In recent years, peanuts have been added to a variety of cereal- and legume-based foods to alleviate the problem of malnutrition. Moreover, peanuts in the form of flour, protein isolate, and meal in a mixed product have desirable sensory qualities (Singh and Singh 1991). Peanut flour is made by crushing the shelled, skinned nuts, extracting the oil, and grinding the crushed nuts. In India, the flour is used in supplementary foods, weaning foods, and protein-rich biscuits (Achaya 1980).

In addition, the peanut plant itself has a high nutritional value and can be used for livestock feed or plowed back into the soil to aid in fertilization of future crops (Cole and Dorner 1992). Nonedible nuts are processed into oil, with the cake used for animal feed. Peanut shells, which accumulate in abundance, can be used as fuel for boilers (Woodroof 1966).

Nutritional Value

Protein

Having a higher percentage of protein by weight than animal foods and beans (ranging from 22 to 30 percent), peanuts provide an excellent, inexpensive source of vegetable protein for humans. A 1 ounce serving of oil- or dry-roasted peanuts provides 7 to 8 grams of protein, or 11 to 12 percent of the U. S. Recommended Dietary Allowance (RDA).

The protein quality of the peanut also is high, with liberal amounts of most of the essential and nonessential amino acids (the limiting amino acids in roasted peanuts and peanut butter are lysine, threonine, methionine, and cystine). For this reason, U. S. government nutritional guidelines include peanuts along with other high-quality protein foods, such as meat, poultry, fish, dry beans, and eggs. In the last few decades, cereal- and legume-based plant food mixtures using peanuts have grown in popularity, especially in developing countries, because of the excellent nutritional value of peanut proteins and their low cost. Table II.D.2.2 presents some chemical indices of protein quality for peanuts and other high-protein foods.

New methods for determining free amino acids in whole peanuts are now available (Marshall, Shaffer,

Table II.D.2.2. *Comparison of various indexes of protein quality for peanuts and other protein-rich foods*

Protein source	Essential amino acid index	Observed biological value	PER	NPU
Peanuts	69	57	1.65	42.7
Egg, whole	100	96	3.92	93.5
Beef	84	76	2.30	66.9
Fish	80	85	3.55	79.5
Milk, cow	88	90	3.09	81.6
Beans	80	59	1.48	38.4
Soybeans	83	75	2.32	61.4
Wheat	64	67	1.53	40.3

Note: The essential amino acid index rates protein quality with respect to all of the 11 essential amino acids. The PER (protein efficiency ratio) is an animal bioassay that measures the efficiency of a protein in producing weight gain in rats. The NPU (net protein utilization) is a similar method that adjusts with a control fed no protein whatsoever, and measures the changes in body nitrogen between the two dietary groups and a group of animals sacrificed at the beginning of each feeding period. BV (biological value) uses estimates of retained nitrogen from the difference between ingested nitrogen and that accounted for in urine and feces.

Source: Samonds and Hegsted (1977), pp. 69–71.

and Conkerkin 1989); Table II.D.2.3 shows that peanuts have an amino pattern similar to that of high-quality proteins, and Table II.D.2.4 indicates that peanuts are much higher in protein than other staple plants, save for legumes.

Peanut proteins include the large saline-soluble globulins, arachin and conarachin, and the water-soluble albumins. The relative protein content of peanuts may vary with variety, strain, growing area, and climate. Arachin constitutes about 63 percent, and conarachin 33 percent, of the total protein in peanuts. The remaining 4 percent consists of other proteins, including glycoproteins, peanut lectin (agglutinin), alpha beta amylase inhibitor, protease inhibitors, and phospholipase D.

Fat

Depending on the cultivar, the fat content of peanuts ranges from 44 to 56 percent. Over 85 percent of the fat in peanuts is unsaturated; an ounce of peanuts contains 14 grams of fat, of which about a third is polyunsaturated and over half is monounsaturated. More precisely, peanuts have a polyunsaturated to saturated fat ratio of 2:3; a high proportion of their total fat is monounsaturated (49 to 54 percent), and a low percentage (14 to 15 percent) is saturated (McCarthy and Matthews 1984).

Monounsaturated fats help to lower LDL (low density lipoprotein) cholesterol when they replace saturated fats in the diet, and thus can help reduce risks of coronary artery disease that are associated with hyperlipidemia.

Calories, Carbohydrates, and Cholesterol

Well over three-quarters of the calories in peanuts are from fat, with the remainder from protein and carbohydrate, although the content of the latter varies with

variety and growing conditions. Peanuts usually have about 20 percent carbohydrates, most of which are sucrose (4 to 7 percent) and starch (0.5 to 7 percent). Peanuts have no cholesterol.

Table II.D.2.3. *Comparison of the amino acids in peanuts compared to high-quality proteins*

Amino acid	Mg/g protein	
	Peanuts	High-quality protein pattern
Histidine	27	17
Isoleucine	48	42
Methionine and cystine	27	26
Phenylalanine and tyrosine	99	73
Threonine	31	35
Tryptophan	13	11
Valine	58	48
Lysine	41	51

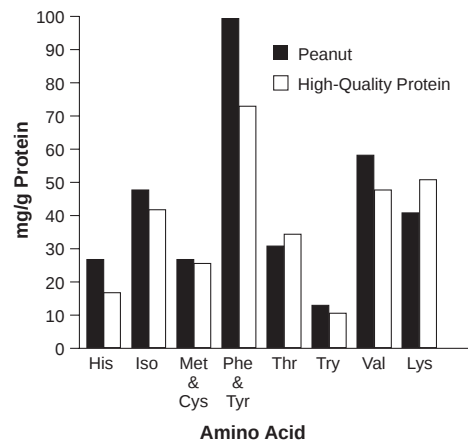


Table II.D.2.4. *Comparison of nutritive value of peanuts with other common cereals and legumes (g/100 g)*

	Water	Protein	Fat	Carbohydrate	Ash	Botanical name
Corn	13.0	8.8	4.0	73.0	1.2	<i>Zea mays</i>
Lentils	11.2	25.0	1.0	59.5	3.3	<i>Lens culinaris</i>
Peanuts	4.0	26.2	42.8	24.3	2.7	<i>Arachis hypogaea</i>
Rice	12.0	7.5	1.7	77.7	1.1	<i>Oryza sativa</i>
Soybeans	7.5	34.9	18.1	34.8	4.7	<i>Glycine soja</i>
Wheat	12.5	12.3	1.8	71.7	1.7	<i>Tritian esculentum</i>

Source: Table based on Spector (1956).

Fiber

The dietary fiber content of peanuts is approximately 7 percent by weight. The percentage of edible fiber is 3.3 (Ockerman 1991a), and of water-soluble fiber 0.77. The latter two percentages were determined by enzymatic figures (Deutsche Forschungsanstalt für Lebensmittelchemie 1991).

Sodium

In their raw state, peanuts are very low in sodium. Unsalted dry-roasted nuts, and “cocktail” nuts, contain no sodium in a 1-ounce serving. However, whole peanuts are usually served salted. A 1-ounce serving of lightly salted peanuts contains less than 140 milligrams of sodium, which is the U. S. Food and Drug Administration’s current definition of a low sodium food. But other peanut products, such as “regular salted” nuts, contain higher amounts of sodium.

Vitamins and Minerals

Peanuts are good sources of riboflavin, thiamine, and niacin, and fair sources of vitamins E and K. They are also relatively high in magnesium, phosphorous, sulfur, copper, and potassium.

In the case of niacin, peanuts are rich sources of tryptophan (an essential amino acid that can be converted into niacin) and, in addition, are relatively rich sources of preformed niacin itself, a 1-ounce serving providing 20 percent of the U.S. RDA.

Nutrients and Processing

Under processing conditions in developing countries, dry roasting preserves both the storage stability of peanuts and their nutritional value to a greater extent than oil roasting (DaGrame, Chaven, and Kadam 1990). With roasting, the thiamine content decreases and the color darkens; hence color gives an indication of the extent of thiamine loss. The proteins, vitamins (except thiamine), and minerals are very stable during processing. But blanching or mechanical removal of skin further reduces thiamine content because thiamine is concentrated in the skins (Woodroof 1966). Table II.D.2.5 shows the nutritional composition of *Arachis hypogaea* L.

Table II.D.2.5. *Nutritional value of Arachis Hypogaea L.*

Constituent	Peanuts, plain (/100g)
Water (g)	6.3
Protein (g)	25.6
Fat (g)	46.1
Carbohydrate (g)	12.5
Energy value (kcal)	564
Total nitrogen (g)	4.17
Fatty acids - saturated (g)	8.2
Fatty acids - monounsaturated (g)	21.1
Fatty acids - polyunsaturated (g)	14.3
Cholesterol (mg)	0
Starch (g)	6.3
Total sugars (g)	6.2
Dietary fiber - southgate method (g)	7.3
Dietary fiber - Englyst method (g)	6.2
Na (mg)	2
K (mg)	670
Ca (mg)	60
Mg (mg)	210
P (mg)	430
Fe (mg)	2.5
Cu (mg)	1.02
Zn (mg)	3.5
Cl (mg)	7
Mn (mg)	2.1
Se (µg)	3
I (µg)	3
Vitamin E (mg)	10.09
Thiamin (mg)	1.14
Riboflavin (mg)	0.10
Niacin (mg)	13.8
Trypt/60 (mg)	5.5
Vitamin B ₆ (mg)	0.59
Folate (µg)	110
Pantothenate (mg)	2.66
Biotin (µg)	72.0
Amino acids (g)	-
Arginine	6.9
Histidine	1.3
Isoleucine	2.6
Leucine	4.1
Lysine	1.9
Methionine	0.6
Phenylalanine	3.1
Threonine	1.6
Tryptophan	0.8
Tyrosine	-
Valine	2.8

Source: Holland et al. (1991) and Ockerman (1991), p. 1331.

Health-Related Issues

Enhancing Protein Quality

Plant proteins, like those in peanuts, which are rich in essential amino acids and nitrogen and low in only a few amino acids, help improve the overall quality of diets, especially diets based on plant proteins. Protein supplementation involves adding to the diet small amounts of a protein that is a rich source of those amino acids that would otherwise be lacking. Protein complementation involves combining protein sources so that they mutually balance each other's excesses or deficiencies (Bressani 1977).

These principles have been used to produce cereal-legume multimixes for humans (Bressani and Elias 1968), and the cuisines of several countries that have traditionally relied on plant protein foods as staples also employ these same principles to good effect, so that protein quality is rarely a problem.

A Weaning Food in Cereal Multimixes

Infants and young children, as weanlings, are growing rapidly and require plenty of high-quality protein. Yet, for cultural and economic reasons, protein-rich animal foods are frequently not readily available in many developing countries. A quarter of a century ago, cereal and cereal-legume multimixes (including peanuts) began to be used to provide a high-protein and high-calorie weaning food for children in this age group. These multimixes can be produced at the local level, are economical, and have excellent results in supporting child growth.

Similarly, many protein-rich cereal- and legume-based foods containing peanuts are now in widespread use in developing countries for alleviating problems associated with protein calorie malnutrition. Peanuts, which are rich in oil and in protein, and are also tasty, are particularly valuable for these purposes (Singh and Singh 1991).

Allergenicity

It is unfortunate that peanuts are not for everyone. The cotyledons, axial germ tissue (hearts), and skin of peanuts contain allergens, and some, but not all, of these are still present after roasting. Because the allergens do not have a characteristic odor or flavor, they cannot easily be detected by peanut-sensitive individuals; thus, labeling of peanut-containing products is essential, save for pure peanut oil, which is not allergenic.

The many different peanut allergens that exist all contain protein. These allergens have been demonstrated by the use of radioallergenabsorbent test (RAST) inhibition, and by immunologic techniques, such as crossed radioimmunoelectrophoresis (CRIE), two-dimensional electrophoresis, and immunoblotting to isolate and characterize the peanut allergens. Then sera from peanut-sensitive individuals is used to determine if specific IgE binding to the various isolated sub-

fractions exists. Since the isolation and characterization methods may affect the physical structure of the protein or its subfractions, different techniques may give different results. Nonetheless, at present, it is clear that multiple allergens exist in peanuts.

A well-characterized peanut allergen recently identified in patients with atopic dermatitis and positive peanut challenges is called Ara h I (*Arachis hypogaea* I) in the official nomenclature (Burks et al. 1991). Highly atopic infants and children appear to be particularly likely to form IgE antibodies that respond to peanuts, as well as to other food proteins (Zimmerman, Forsyth, and Gold 1989). Such children begin producing IgE antibodies to respond to inhalant allergens during their first and second years of life; they are defined as highly atopic because their serum IgE levels are 10 times those of normal infants, and their RAST tests are positive on multiple occasions.

Diagnosis. Diagnosis of peanut allergy is difficult because standardized peanut extracts do not yet exist. A RAST can be used on the sera of already sensitive persons for whom a skin test would be dangerous. Double-blind placebo-controlled challenges are definitive but are not often needed. If such double-blind challenges are done, provisions need to be made to cope with emergencies that may arise if an anaphylactic reaction occurs. The allergenicity of hydrolyzed peanut protein must be further studied. It is not clear at what level of hydrolysis allergenicity is lost.

Prevalence of peanut sensitivity. Peanut sensitivity is less prevalent among humans than are, for example, sensitivities to milk and eggs, but peanuts are especially dangerous for a number of reasons. One important reason is that peanuts occur in small amounts in so many different foods and recipes, ranging from satay sauce and "vegeburgers" to main dishes and spreads, making it difficult for those who have the allergy to avoid them (Smith 1990).

The allergy occurs in vegetarians as well as omnivores (Donovan and Peters 1990), and cross-reactivity with other legumes appears to exist. In addition, individuals who are allergic to other foods besides legumes are sometimes allergic to peanuts. Finally, highly atopic individuals, such as asthmatics, and those who suffer from atopic dermatitis or from multiple other food allergies, are likely to be at particular risk.

Signs of peanut allergy. The signs of peanut sensitivity range from urticaria (hives) to angioedema and asthma, and occasionally even to anaphylaxis and death (Lemanske and Taylor 1987; Boyd 1989). Crude extracts of proteins in both raw and roasted peanuts, as well as purified peanut proteins, such as arachin, conarachin, and concanavalin A reactive glycoprotein, are all allergenic in some persons (Barnett, Baldo, and Howden 1983).

Natural history of peanut allergy. Allergic reactions to peanuts usually begin early in life and persist. Studies over a period of several years have now been completed on individuals who exhibited symptoms of peanut allergy in childhood after a double-blind, placebo-controlled challenge, and reacted positively to puncture skin tests at the same time. Most such individuals had avoided peanuts since diagnosis, but those who inadvertently ingested peanuts 2 to 14 years later had reactions. This, coupled with a continued skin reactivity to peanut extract in puncture tests, suggests that peanut-sensitive individuals rarely lose their sensitivity with time (Bock and Atkins 1989). Fatal reactions to peanuts can also occur after many years of abstinence (Fries 1982). Fortunately, peanut oil, at least the usual grade sold in the United States and Europe, which contains no detectable protein, is not allergenic (Taylor et al. 1982). Unfortunately, in other countries the oil may contain enough of the protein to cause allergic reaction.

Allergy treatment. Avoidance of products containing peanut protein is the surest way to avoid peanut allergy. But certain foods may be accidentally contaminated with peanut protein, so that even products supposedly peanut free may be dangerous. Because the prevalence of peanut allergy is high, both labeling and label reading are important. Treatment of peanut sensitivity with immunotherapy has not proved helpful. If a sensitive person does ingest peanuts, self-administered epinephrine may help.

Peanut anaphylaxis is a medical emergency (Sampson 1990). One common cause is consumption of a product containing decaffeinated and colored peanut protein reformulated to resemble other nuts (Yunginger et al. 1989). In one case, an "almond" icing that was actually made from peanuts led to a fatal reaction (Evans, Skea, and Dolovich 1988). Other possible hidden sources of peanuts are egg rolls, cookies, candy, pastries, and vegetable burgers. Chinese food and dried-food dressings that contain peanuts have also been causes of anaphylactic shock (Assem et al. 1990).

Peanut allergy is probably the major cause of food-related anaphylaxis in the United States. Only a few milligrams will cause reactions in some persons. Those who are at risk of anaphylactic reactions to peanuts should wear medic-alert bracelets and carry preloaded epinephrine syringes and antihistamines. If treatment is needed, repeated doses of epinephrine, antihistamines, corticosteroids, mechanical methods to open airways, oxygen, vasopressors, and intravenous fluids may be necessary to prevent a fatal reaction (Settipane 1989).

Cross-reactivity in allergenicity. Peanuts cross-react in vitro with other members of the Leguminosae family, especially with garden peas, chickpeas, and soybeans, although clinical sensitivity is not always

observed (Toorenenbergen and Dieges 1984; Barnett, Bonham, and Howden 1987). Reactions to nonlegume nuts, however, are relatively rare among those allergic to peanuts.

Aflatoxin

Aflatoxins are naturally occurring environmental contaminants that often infest peanuts. These carcinogenic mycotoxins arise from a fungus (*Aspergillus flavus*) that colonizes peanuts under certain environmental conditions or improper storage conditions that permit fungal growth. The aflatoxins B₁ and B₂ are most common in peanuts.

Aflatoxin-contaminated food has been shown to be associated with liver cancer in both humans and several experimental animals. The problem appears to be most severe in Africa and other parts of the world where environmental and storage conditions favor the mold's growth and where processing methods that could identify and eliminate contaminated seeds are still not in place. But it is also a problem in the Orient, where people prefer the flavor of crude peanut oil. This form of the oil is high in aflatoxins.

In highly industrialized countries, aflatoxin contamination generally occurs before the harvest; in developing countries, contamination during storage is an additional and seemingly greater problem. In the former case, insect damage to the pods and roots permits seed contamination by the mold, especially during growing seasons where there is a late-season drought, which increases stress on the plant and also helps to exclude competition by other fungi. In a three-year study it was found that pods are more susceptible to contamination than roots (Sanders et al. 1993). Shriveled peanuts have the highest content of aflatoxin B₁ (Sashidhar 1993).

Under favorable environmental conditions, *Aspergillus flavus* produces millions of spores, which can be carried by the wind for miles (Cleveland and Bhatnagar 1992). Insect damage promotes *Aspergillus flavus* growth when the spores land at a site of insect injury and when the moisture content exceeds 9 percent in peanuts, or 16 percent in peanut meal, and temperatures are 30 to 35° C. Therefore, the only sure method to avoid aflatoxins is to prevent their formation by careful harvesting and quick drying and storage. Food additives, such as sodium bisulfite, sorbate, propionate, and nitrite, reduce aflatoxin production. Also, certain food components and spices, such as peppers, mustard, cinnamon, and cloves, may inhibit mycotoxin production (Jones 1992).

Visual screening of peanuts will reveal the conical heads of the *Aspergillus flavus* fungus. Another technique is to screen unshelled nuts for the presence of bright greenish yellow fluorescence (BGYF) under ultraviolet light, using electronic color-sorting techniques. However, neither of these techniques screens out all aflatoxin contamination. For greater precision, chemical tests are used that include thin

layer chromatography (TLC) and high performance liquid chromatography (HPLC) (Beaver 1989).

Immunological methods, such as ELISA or affinity column methods, are also useful, although precautions must be taken in performing the analysis since aflatoxins are highly carcinogenic (Wilson 1989). Much research is now in progress on the development of standardized methods for determining aflatoxin in peanut products. Other research concentrates on eliminating the contamination or inactivating it.

Chlorine gas can inactivate one aflatoxin, aflatoxin B₁, and the resulting compounds do not appear to be mutagenic (Samarajeewa et al. 1991). Ammonia and ozone treatments of peanuts also appear to work. As yet, however, these methods are still experimental.

The liver cancer in question is thought to be caused by hepatitis B, with other factors, such as aflatoxin, acting as promoters or as less-potent initiators. In most case-control studies, primary hepatocellular carcinoma is highly associated with antibodies to hepatitis B and C virus; other risk factors, such as peanut consumption (presumably with aflatoxin contamination), smoking, and drinking, are less highly associated (Yu et al. 1991). A recent large study in mainland China showed high correlations of liver cancer with hepatitis B surface antigen (HBsAg+) carrier status, whereas lesser associations were seen with alcohol use and cadmium of plant origin, but none with the measure of aflatoxin exposure that was used (Campbell et al. 1990). However, smaller studies, especially of peanut oils contaminated with aflatoxin, continue to suggest that aflatoxin may be involved in hepatobiliary cancers (Guo 1991).

Aflatoxin: Food safety. In recent years, the European Community has increasingly collaborated on food safety tests, and a great deal of effort has been devoted to developing more sensitive, rapid, and standardized methods for detecting aflatoxin levels in peanut products (Van Egmond and Wagstaffe 1989; Patey, Sherman, and Gilbert 1990; Gilbert et al. 1991). Thanks to this collaboration, standardized methods and materials are now available for assuring valid and reliable food safety testing. Current guidelines are that no more than 20 parts per billion of aflatoxin are permitted.

Developing countries and the aflatoxin problem. As we have mentioned, the problem of aflatoxin contamination is particularly serious in developing countries because of storage and processing problems. Thus, by way of example, aflatoxin levels in Pacific countries such as Fiji and Tonga are high, and so are liver cancer rates. Therefore, in addition to improving inspection methods, techniques must be developed for decreasing the carcinogenicity of aflatoxin-contaminated foods for humans and for animals. One technique that has been helpful in reducing the carcinogenicity of aflatoxin-contaminated groundnut cakes is ammoniation (Frayssinet and Lafarge-Frayssinet 1990). However, more such solutions are needed.

Table II.D.2.6. *Nutritive value of common peanut foods (g/100g)*

	Nut, roasted (Spanish or Virginia, salted)	Peanut butter	Peanut oil
Water, g	1.6	1.7	0
Calories	586	587	884
Protein, g	26	25	0
Carbohydrate, g	19	19	0
Fat	50	51	100

Miscellaneous

Because peanuts are a low-acid food, incorrect commercial canning of peanuts can cause botulism, as a recent case in China showed (Tsai et al. 1990). On a more positive note, peanuts are often used by dentists as foods for masticatory tests for jaw and muscle function (Kapur, Garrett, and Fischer 1990), because they are easy to standardize and familiar to patients (Wang and Stohler 1991). Moreover, roasted peanuts are among the least likely of popular snack foods to cause dental caries (Grenby 1990). Peanuts remain available dry, oil roasted, unsalted, lightly salted, or salted, as well as in a variety of flavors today in the United States (Nabisco Foods Group, Planters Division 1990). And finally, the not fully mature (or green) peanuts can be used as a vegetable that is much appreciated in the southern United States. Green peanuts are boiled in a weak brine solution and usually consumed immediately. If refrigerated, they will last for five days. They can also be frozen or canned.

Table II.D.2.6 shows the nutritive value of various peanut foods.

Johanna T. Dwyer and Ritu Sandhu

Bibliography

- Achaya, K. T. 1980. Fat intakes in India. In *Combating under-nutrition*, ed. C. Gopalan, 110-12. New Delhi.
- Assem, E. S., G. M. Gelder, S. G. Spiro, et al. 1990. Anaphylaxis induced by peanuts. *British Medical Journal* 300: 1377-8.
- Barnett, D., B. A. Baldo, and M. E. Howden. 1983. Multiplicity of allergens in peanuts. *Journal of Allergy and Clinical Immunology* 72: 61-8.
- Barnett, D., B. Bonham, and M. E. H. Howden. 1987. Allergic cross reactions among legume foods: An in vitro study. *Journal of Allergy and Clinical Immunology* 79: 433-8.
- Basha, S. M. 1991. Accumulation pattern of arachin in maturing peanut seeds. *Peanut Science* 16: 70-3.
- Beaver, R. W. 1989. Determination of aflatoxins in corn and peanuts using high performance liquid chromatography. *Archives of Environment Contamination and Toxicology* 18: 315-18.
- Bock, S. A., and F. M. Atkins. 1989. The natural history of

- peanut allergy. *Journal of Allergy and Clinical Immunology* 83: 900-4.
- Boyd, G. K. 1989. Fatal nut anaphylaxis in a 16-year-old male: Case report. *Allergy Proceedings* 10: 255-7.
- Branch, W. D., and C. K. Kvien. 1992. Peanut breeding for drought resistance. *Peanut Science* 19: 44-6.
- Bressani, R. 1977. Protein supplementation and complementation. In *Evaluation of proteins for humans*, ed. C. E. Bodwell, 204-31. Westport, Conn.
- Bressani, R., and L. G. Elias. 1968. Processed vegetable protein mixtures for human consumption in developing countries. In *Advances in food research*, Vol. 16, ed. E. W. Mrak, G. F. Stewart, and C. O. Chichester, 1-78. New York.
- Burks, A. W., L. W. Williams, R. M. Helm, et al. 1991. Identification of a major peanut allergen Ara h I in patients with atopic dermatitis and positive peanut challenges. *Journal of Allergy and Clinical Immunology* 88: 172-9.
- Burks, A. W., L. W. Williams, S. B. Mallory, et al. 1989. Peanut protein as a major cause of adverse food reactions in patients with atopic dermatitis. *Allergy Proceedings* 10: 265-9.
- Bush, R. K., S. L. Taylor, and J. A. Nordlee. 1989. Peanut sensitivity. *Allergy Proceedings* 10: 13-20.
- Campbell, T. C., J. S. Chen, C. B. Liu, et al. 1990. Nonassociation of aflatoxin with primary liver cancer in a cross-sectional ecological survey in the People's Republic of China. *Cancer Research* 50: 6882-93.
- Castelli, W. P. 1992. Concerning the possibility of a nut. . . . *Archives of Internal Medicine* 152: 1371-2.
- Cleveland, T. E., and D. Bhatnagar. 1992. Aflatoxin: Elimination through biotechnology. In *Encyclopedia of food science and technology*, Vol. 1, ed. Y. H. Hui, 6-11. New York.
- Cole, R. J., and J. W. Dornier. 1992. Peanuts. In *Encyclopedia of food science and technology*, Vol. 3, ed. Y. H. Hui, 2036-9. New York.
- DaGrame, S. V., J. K. Chavan, and S. D. Kadam. 1990. Effects of roasting and storage on proteins and oil in peanut kernels. *Plant Foods for Human Nutrition* 40: 143-8.
- Deutsche Forschungsanstalt für Lebensmittelchemie. 1991. *Food composition and nutrition tables 1989-90*. Stuttgart.
- Donovan, K. L., and J. Peters. 1990. Vegetable burger allergy: All was nut as it appeared. *British Medical Journal* 300: 1378.
- Elin, R. J., and J. M. Hosseini. 1993. Is magnesium content of nuts a factor in coronary heart disease? *Archives of Internal Medicine* 153: 779-80.
- Evans, S., D. Skea, and J. Dolovich. 1988. Fatal reaction to peanut antigen in almond icing. *Canadian Medical Association Journal* 193: 231-2.
- FAO (Food and Agriculture Organization of the United Nations). 1970. *Amino-acid content of foods*. Rome.
1993. *Food and Agriculture Organization Yearbook: Production, 1992*, Vol. 46. Rome.
- Food and Nutrition Board, National Research Council. 1989. *Recommended dietary allowances*. Tenth edition. Washington, D.C.
- Fraser, G. E., J. Sabate, W. L. Beeson, and T. M. Strahan. 1992. A possible protective effect of nut consumption on risk of coronary heart disease. *Archives in Internal Medicine* 152: 1416-26.
- Frayssinet, C., and C. Lafarge-Frayssinet. 1990. Effect of ammoniation on the carcinogenicity of aflatoxin-contaminated groundnut oil cakes: Long-term feeding study in the rat. *Food Additives and Contaminants* 7: 63-8.
- Fries, J. H. 1982. Peanuts: Allergic and other untoward reactions. *Annals of Allergy* 48: 220-6.
- Gilbert, J., M. Sharman, G. M. Wood, et al. 1991. The preparation, validation and certification of the aflatoxin content of two peanut butter reference materials. *Food Additives and Contaminants* 8: 305-20.
- Greene, K. S., J. M. Johnson, M. Rossi, et al. 1991. Effects of peanut butter on ruminating. *American Journal on Mental Retardation* 95: 631-45.
- Grenby, T. H. 1990. Snack foods and dental caries: Investigations using laboratory animals. *British Dental Journal* 168: 353-61.
- Guo, H. W. 1991. The study of the relationship between diet and primary liver cancer. *Chinese Medical Journal* 25: 342-4.
- Holaday, C. E., and J. L. Pearson. 1974. Effect of genotype and production area on the fatty acid composition, total oil and total protein in peanuts. *Journal of Food Science* 39: 1206-9.
- Holbrook, C. C., and J. P. Noe. 1992. Resistance to the peanut root-knot nematode (*Meloidogyne arenaria*) in *Arachis hypogaea*. *Peanut Science* 19: 35-7.
- Holland, B., A. A. Welch, I. D. Unwin, et al. 1991. *McCance and Widdowson's - The composition of foods*. Fifth edition. London.
- Jambunathan, G., S. Gurtu, K. Raghunath, et al. 1992. Chemical composition and protein quality of newly released groundnut (*Arachis hypogaea* L.) cultivars. *The Science of Food and Agriculture* 59: 161-7.
- Johnson, P., and W. E. F. Naismith. 1953. The physicochemical examination of the conarachin fraction of the groundnut globulins (*Arachis hypogaea*). *Discussions of the Faraday Society* 13: 98.
- Jones, J. 1992. *Food safety*. New York.
- Kannon, G., and H. K. Park. 1990. Utility of peanut agglutinin (PNA) in the diagnosis of squamous cell carcinoma and keratoacanthoma. *American Journal of Dermatopathology* 12: 31-6.
- Kapur, K. K., N. R. Garrett, and E. Fischer. 1990. Effects of anaesthesia of human oral structures on masticatory performance and food particle size distribution. *Archives of Oral Biology* 35: 397-403.
- Klevay, L. M. 1993. Copper in nuts may lower heart disease risk. *Archives of Internal Medicine* 153: 401-2.
- Krapovickas, A. 1969. The origin, variability and spread of the groundnut (*Arachis hypogaea*). In *Domestication and exploitation of plants and animals*, ed. P. J. Ucko and G. W. Dimbleby, 427-41. Chicago.
- Langkilde, N. S., H. Wolf, H. Clausen, and T. F. Orntoft. 1992. Human urinary bladder carcinoma glycoconjugates expressing T-(Gal-beta [1-3] GalNAc-alpha-I-O-R) and T-like antigens: A comparative study using peanut agglutinin and poly and monoclonal antibodies. *Cancer Research* 52: 5030-6.
- Lapidis, D. N., ed. 1977. *Encyclopedia of food agriculture and nutrition*. New York.
- Lemanske, R. F., and S. L. Taylor. 1987. Standardized extracts of foods. *Clinical Reviews in Allergy* 5: 23-6.
- Marshall, H. F., G. P. Shaffer, and E. J. Conkerkin. 1989. Free amino acid determination in whole peanut seeds. *Analytical Biochemistry* 180: 264-8.
- McCarthy, M. A., and R. H. Matthews. 1984. *Composition of foods: Nuts and seed products*. Washington, D.C.
- McGee, Harold. 1988. *On food and cooking - The science and lore of the kitchen*. London.
- Nabisco Foods Group, Planters Division. 1990. *Why everybody loves a peanut*. Winston-Salem, N.C.

- Ockerman, H. W. 1991a. *Food science sourcebook: Food composition, properties and general data*, Vol. 2. New York.
- 1991b. *Food science sourcebook: Terms and descriptions*, Vol. 1. New York.
- Oohira, H. 1992. Morphological and histochemical studies on experimentally immobilized rabbit patellar cartilage. *Medical Journal of Kagoshima University* 44: 183-223.
- Oser, B. L. 1959. An integrated essential amino acid index for predicting the biological value of proteins. In *Protein and amino acid nutrition*, ed. A. A. Albanese, 292-5. New York.
- Patey, A. L., M. Sherman, and J. Gilbert. 1990. Determination of aflatoxin B1 levels in peanut butter using an immunoaffinity column clean-up procedure: Inter-laboratory study. *Food Additives and Contaminants* 7: 515-20.
- Porter, D. M., T. A. Coffelt, F. S. Wright, and R. W. Mazingo. 1992. Resistance to sclerotinia blight and early leaf spot in Chinese peanut germplasm. *Peanut Science* 19: 41-3.
- Sabate, J., G. E. Fraser, K. Burke, et al. 1993. Effects of walnuts on serum lipid levels and blood pressure in normal men. *New England Journal of Medicine* 328: 603-7.
- Sachs, M. I., R. T. Jones, and J. W. Yunginger. 1981. Isolation and partial characterization of a major peanut allergen. *Journal of Allergy and Clinical Immunology* 67: 27-34.
- Samarajeewa, U., A. C. Sen, S. Y. Fernando, et al. 1991. Inactivation of aflatoxin B1 in cornmeal, copra meal and peanuts by chlorine gas treatment. *Food and Chemical Toxicology* 29: 41-7.
- Samonds, K. W., and D. M. Hegsted. 1977. Animal bioassays: A critical evaluation with specific reference to assessing the nutritive value for the human. In *Evaluation of proteins for humans*, ed. C. E. Bodwell, 69-71. Westport, Conn.
- Sampson, H. A. 1990. Peanut anaphylaxis. *Journal of Allergy and Clinical Immunology* 86: 1-3.
- Sanders, T. H., R. J. Cole, P. D. Blankenship, and J. W. Dörner. 1993. Aflatoxin contamination of the peanuts from plants drought stressed in the pod or root zones. *Peanut Science* 20: 5-8.
- Sauer, J. D. 1993. *Historical geography of crop plants - A select roster*. Boca Raton, Fla.
- Scott, P. M. 1969. The analysis of food for aflatoxin and other fungal toxins - A review. *Canadian Institute of Food Science and Technology Journal* 2: 173-7.
- Senba, M., K. Watanabe, K. Yoshida, et al. 1992. Endocarditis caused by *Candida parapsilosis*. *Southeast Asian Journal of Tropical Medicine and Public Health* 23: 138-41.
- Settipane, G. A. 1989. Anaphylactic deaths in asthmatic patients. *Allergy Proceedings* 10: 271-4.
- Sashidhar, R. B. 1993. Fate of aflatoxin B-1 during the industrial production of edible defatted peanut protein flour from raw peanuts. *Food Chemistry* 48: 349-52.
- Short, J. 1990. *Grain legumes: Evolution and genetic resources*. New York.
- Singh, B., and H. Singh. 1991. Peanut as a source of protein for human foods. *Plant Foods for Human Nutrition* 41: 165-77.
- Smith, T. 1990. Allergy to peanuts. *British Medical Journal* 300: 1354.
- Spector, W. S., ed. 1956. *Handbook of biological data*. Bethesda, Md.
- Taylor, S. L., W. W. Busse, M. I. Sachs, et al. 1982. Peanut oil is not allergenic in peanut-sensitive individuals. *Journal of Allergy and Clinical Immunology* 68: 372-5.
- Toorenenbergen, A. W. van, and P. H. Dieges. 1984. IgE mediated hypersensitivity to taugeh (sprouted small green beans). *Annals of Allergy* 53: 239-42.
- Tsai, S. J., Y. C. Chang, J. D. Wong, and J. H. Chon. 1990. Outbreak of type A botulism caused by a commercial food product in Taiwan: Chemical and epidemiological investigations. *Chinese Medical Journal* 46: 43-8.
- USDA (United States Department of Agriculture). 1992. *USDA - Agriculture statistics, 1992*. Washington, D.C.
- Van Egmond, H. P., and P. J. Wagstaffe. 1989. Aflatoxin B1 in peanut meal reference materials: Intercomparisons of methods. *Food Additives and Contaminants* 6: 307-19.
- Wang, J. S., and C. S. Stohler. 1991. Predicting food stuff from jaw dynamics during masticatory crushing in man. *Archives of Oral Biology* 36: 239-44.
- Weijian, Z., P. Shiyao, and L. Mushon. 1991. Changes in the morphology and structure of cotyledon storage cells and their relation to the accumulation of oil and protein in the peanut. *Scientia Agricultura Sinica* 24: 8-13.
- Willich, R. K., N. J. Morris, and A. F. Freeman. 1954. The effect of processing and storage of peanut butter on the stability of their oils. *Food Technology* 8: 101-4.
- Wilson, D. M. 1989. Analytical methods for aflatoxin in corn and peanuts. *Archives of Environmental Contamination and Toxicology* 18: 308-14.
- Woodroof, J. G. 1966. *Peanuts: Production, processing, products*. Westport, Conn.
- Young, V. R., and N. S. Scrimshaw. 1977. Human protein and amino acid metabolism and requirements in relation to protein quality. In *Evaluation of proteins for humans*, ed. C. E. Bodwell, 11-54. Westport, Conn.
- Yu, M. W., S. L. You, A. S. Chang, et al. 1991. Association between hepatitis C virus antibodies and hepatocellular carcinoma in Taiwan. *Cancer Research* 51: 5621-5.
- Yunginger, J. W., D. L. Squillace, R. T. Jones, and R. M. Helm. 1989. Fatal anaphylactic reactions induced by peanuts. *Allergy Proceedings* 10: 249-53.
- Zimmerman, B., S. Forsyth, and M. Gold. 1989. Highly atopic children: Formation of IgE antibody to food protein, especially peanut. *Journal of Allergy and Clinical Immunology* 83: 764-70.



II.E

Animal, Marine, and Vegetable Oils

II.E.1 An Overview of Oils and Fats, with a Special Emphasis on Olive Oil

Oils from vegetable sources are playing an increasingly important role in human nutrition, and they, along with foods incorporating them, currently compose 30 percent of the calories in a typical Western diet. In the past, however, vegetable oil utilization was limited, and animal and marine fats were far more important. This chapter discusses the nutritional value of fats and oils widely used in the world today.

Fatty Acid Nomenclature

Most oils consist primarily of triacylglycerols, which are composed of 3 fatty acids esterified to a glycerol molecule. The “essential” fatty acids are a major consideration in assessing the nutritional value of fats and oils. G. O. Burr and M. M. Burr first discovered that fats and oils contained substances that were essential for normal growth and reproduction (Holman 1992), and in the 1970s, convincing evidence was presented to show that the omega-3 fatty acids were also essential in humans. In addition to fatty acids, other components present in fats and oils include fat-soluble vitamins (A, D, and E) and sterols, and as few other compounds in fats and oils are of nutritional importance, only fatty acids, fat-soluble vitamins, and sterols are discussed in this chapter.

The sterols of importance to nutrition include cholesterol and the plant sterols. The structures of common sterols are shown in Figure II.E.1.1. Recently, there has been a great deal of interest in cholesterol oxides, which are formed in cholesterol-containing foods and oils after heating or oxidation and which have been implicated in the development of atherosclerosis (Smith 1987; Nawar et al. 1991).

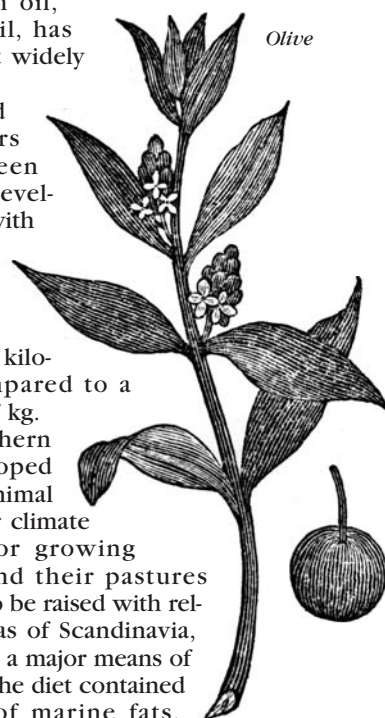
Fatty acid composition is often presented as an average value representing a middle point in the composition of an oil. There are many factors that affect

the fatty acid composition of oilseeds, marine oils, and animal fats. Thus, it is preferable to present a range of fatty acid compositions that are common, rather than to present one value.

World Fat and Oil Production

Current production estimates for important fats and oils are shown in Figure II.E.1.2. Soybean oil is by far the leader, followed by palm, sunflower, and rapeseed oil, with others trailing far behind. The current levels of production of the various vegetable oils reflect relatively recent changes. Palm oil production has increased dramatically in the last few decades, as Malaysia and other Asian countries have promoted it, and in the past half century, soybean oil, once a minor oil, has become the most widely used of all.

Today, fat and oil usage differs markedly between developed and developing countries, with the per capita consumption of fats and oils in industrialized nations about 26 kilograms (kg), compared to a world average of 7 kg. Historically, northern Europeans developed diets based on animal fats because their climate was unsuited for growing oilseed crops and their pastures allowed animals to be raised with relative ease. In areas of Scandinavia, where fishing was a major means of food production, the diet contained large amounts of marine fats.



Olive

In Africa, Asia, and the Mediterranean, however, because the raising of large numbers of animals was impractical and fish were not dietary mainstays, fats were derived primarily from plants, such as oil palms and olive trees. Thus, historical differences in patterns of fat consumption largely resulted from environmental conditions that fostered the various agricultural practices of the different areas of the world.

Yet, over the past 100 years, there has been a decreased consumption of animal fats and an increase in that of vegetable oils. When vegetable oils were first produced in commercially important quantities, they could not be used to displace animal fats because the latter (for example, butter, lard, and tallow) were solid at room temperature and had specific applications that required solidity. But the process of hydrogenation, developed early in the twentieth century, permitted the manufacture of plastic fats (such as shortening and margarine) based on vegetable oils, and lard and butter were subsequently displaced in the diets of many peoples.

The twentieth-century industrialization of oilseed production, as well as vegetable oil extraction and refining, has been the primary factor leading to large-scale changes in the fat composition of diets around the world, although genetic engineering, breeding, and mutation have also brought significant changes in the fatty acid composition of a number of oilseeds. Moreover, the fatty acid composition of animal fats and aquacultured fish can be altered by dietary manipulation. Such changes complicate discussion of the nutritional benefits obtained from oils.

Oil Extraction and Processing

A knowledge of the ways in which oils are obtained (extracted) and processed is important to an understanding of the nutritional value of the oils. Fats and oils are extracted from animal and vegetable materials by the three main processes of rendering, using the screw press, and solvent extraction.

The first method humans employed to extract oils was undoubtedly rendering, which is still used today to remove fat from animal and fish tissues. In this process, the material to be rendered is heated (either dry or in water), which disrupts the tissues and allows separation of the oil. The quality of the oil yielded by early rendering operations must have been poor, but later developments in processing (such as steam heating) have permitted the rendering of high-quality products. The conventional wet-rendering process includes an initial cooking step, in which the tissue is heated with direct or indirect steam, or is heated as it passes through a conveyor. The cooked material is pressed in continuous or batch presses, and the liquid ("press liquor") is centrifuged to separate the water and oil. The oil obtained may be dried further before storage.

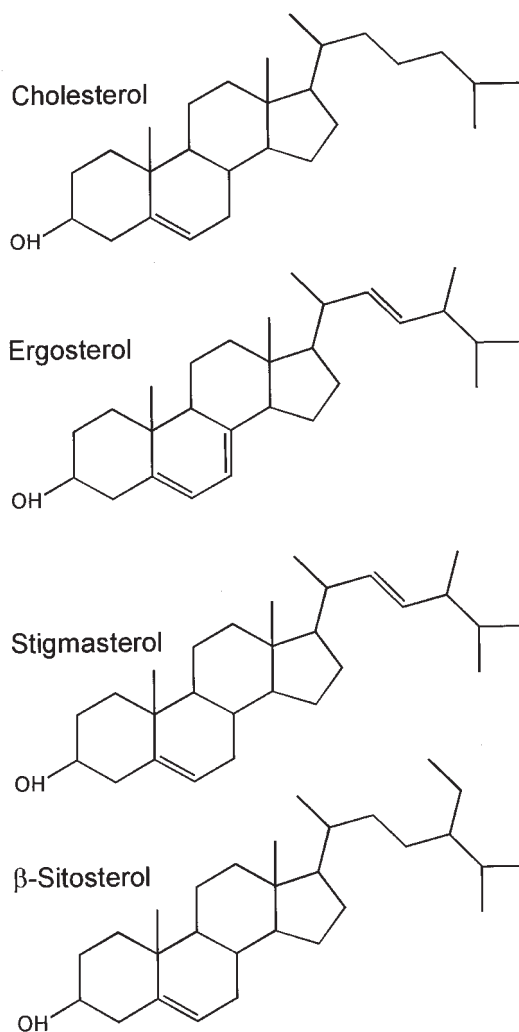


Figure II.E.1.1. The structure of common sterols.

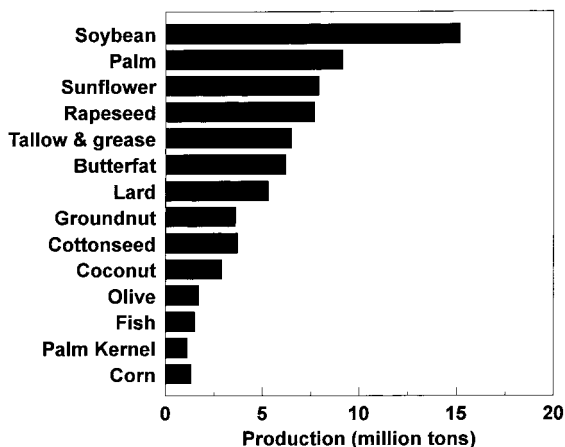


Figure II.E.1.2. Production estimates for important fats and oils.

Table II.E.1.1. Average oil content of plant sources of oil

Material	Oil percent
Soy	18-20
Palm fruit	45-50
Sunflower	25-45
Rapeseed (canola)	35-60
Peanut	45-55
Cottonseed	18-20
Coconut (copra)	63-68
Olive	25-30

In the second method, an expeller, or screw press, removes oil from vegetable materials by mechanical pressure. The pressure creates heat, disrupting the tissue and causing the oil to separate, after which it flows out through holes in the press. Such presses have been documented in early Hindu medical texts; the first of these devices, however, doubtless had low output, and the extracted oil was probably used for illumination or for medicine. These were batch presses, which had to be filled with oilseed and then emptied in each extraction step. By contrast, modern screw presses operate continuously and have much greater capabilities than batch presses. Screw-press extraction of oil is most feasible when the oil content of the material to be pressed is high. The oil content of some vegetable sources of oil is shown in Table II.E.1.1 (Sonntag 1979a).

The third method of oil recovery, solvent extraction, was only possible after supplies of solvents with appropriate characteristics became available. The solvent-extraction process for soybeans is outlined in Figure II.E.1.3. The refining steps include alkali neutralization, which results in the removal of phospholipids and free fatty acids; bleaching, which removes pigments, metals, and free fatty acids; and deodorization, which removes odorous compounds, some sterols, and tocopherols.

W. Normann's discovery of liquid phase hydrogenation (which results in loss of polyunsaturated fatty acids and production of monounsaturated "trans" acids) led to the development of plastic fats derived from vegetable sources. Crisco, the first shortening to be sold (1911), was based on hydrogenated cottonseed oil. By the mid-1940s, 65 percent of the cottonseed oil produced in the United States was used to make shortening.

Vegetable Oils

Olive Oil

Olive oil is derived from the fruit of the evergreen tree *Olea europaea*, which grows in temperate climates with warm and dry summers. Around 5,000 or 6,000 years ago, at the eastern end of the Mediterranean, the tough, spiny, wild olive trees dominating the countryside of Palestine, Syria, and other areas of

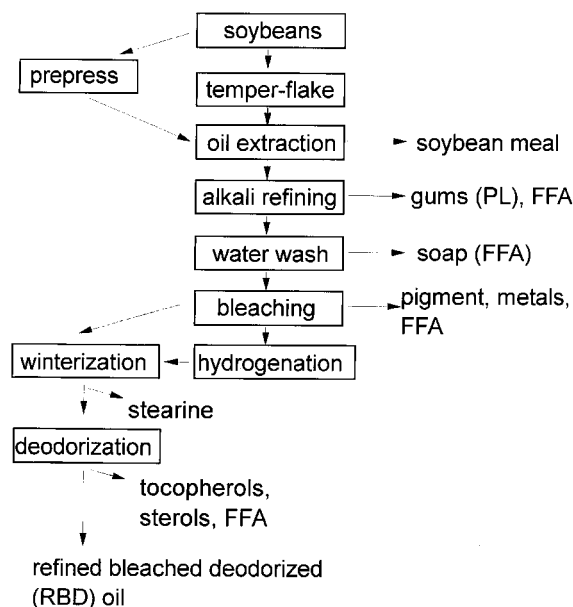


Figure II.E.1.3. Operations in soybean oil extraction and refining.

the Middle East were first brought under cultivation (Chandler 1950). The trees became gnarled with domestication, and less bushy, and their fruit (green that ripens to brown or to blue-purple-black) was laden with oil. That oil – technically a fruit oil rather than a vegetable oil, as it is classified – doubtless was used to fuel lamps, and it found its way into medicines and cosmetics as well as cooking. The many uses for oil created much demand in the ancient world, which was satisfied by oils made from walnuts, almonds, and the seeds of sesame, flax, and radishes, as well as from olives (Tannahill 1988). The latter, however, were the most productive source of oil from the Bronze Age onward, and their cultivation spread throughout the Mediterranean region, so that the waning centuries of that age found the people of the island of Crete cultivating great numbers of olive trees and growing rich on the export of oil (Trager 1995).

Shortly after the dawn of the Iron Age, the Greek landscape reflected a similar dedication to olive production, spurred in the early sixth century B.C. by the prohibition of Solon (the Greek statesman and legal reformer) of the export of any agricultural produce except olive oil. Greece (like Crete, two millennia earlier) would learn how devastating the effects of monoculture could be when war made trade impossible (Tannahill 1988).

Nevertheless, as a relatively precious product requiring special climate and skills, yet easy to store and ship in jars, olive oil lent itself well to this kind of specialization. Barbarians in contact with the Greeks became good customers, although any possibility of

Greek monopoly ended when Etruscans carried the techniques of olive cultivation into Italy, and Phoenicians did the same for North Africa and the Iberian peninsula.

Because many years are required for olive trees to become productive after planting, olive cultivation in these new areas probably required the passing of a generation or two before its possibilities were appreciated. But certainly one thing worth appreciating was that olive trees seemed to live forever. In Spain, for example, it is said that there are trees some 1,000 years old, which meant that in the absence of some catastrophe, such as disease or fire, the task of planting olive trees and then waiting for them to become productive seldom troubled grove owners more than once, if that (Chandler 1950).

In the first century A.D., Pliny noted a dozen varieties of olives grown as far away from Rome as Gaul (France) and Spain, and certainly it was the case that by the time of the Romans, olive-tree cultivation had caught on in Iberia in a big way. With olive oil a staple in the Roman diet, Spanish oil was a commodity sought throughout the Empire, and henceforth olive oil would be established as the most important dietary oil in southern Europe. In the north – where livestock did well and olive trees did not – cooking was done with butter and lard. Obviously, the cuisines these cooking mediums helped to shape were often strikingly dissimilar.

Within fourteenth-century Spain, olive oil was exported from south to north, and wool and hides from north to south, in a kind of microcosm of greater Europe, and, following the conquest of the New World, American demand continued to stimulate Spain's olive oil industry. As the colonists began to make their own oil, however, the flow of Spanish olive oil across the Atlantic diminished. In North America, Thomas Jefferson tried to grow olives at Monticello, but the cuttings he used would not take root (Trager 1995). At about the same time as this failed experiment in the east of the continent, olive cultivation was successfully launched in California by Spanish missionaries, and by the beginning of the twentieth century, that western state had joined Provence in France and the Lucca district in Italy – in the eyes of at least one American writer – as a producer of some of the world's best olive oils (Ward 1911).

Today, 90 percent of the world's olives go into oil, and only 2 percent of the acreage given over to olive production is outside of the Mediterranean region (McGee 1984). In terms of volume, Spain, Italy, Greece, and Portugal are the largest producers, although much of their olive oil is not reflected in production figures because there are many people – with just a few trees – who pick their own olives and take them to local cooperatives to be pressed.

Olives are sometimes picked by hand because they, and the twigs they grow upon, are fragile. How-

ever, because the fruit should be all of the same size and degree of ripeness, the same tree might be picked several times, all of which significantly increases the price of the oil (Toussaint-Samat 1992). To avoid picking several times, the fruits are generally treated less elegantly and are knocked off the trees, either from ground level with long poles or with small rakes by harvesters who climb the trees to shower the olives down. The fruits are caught on cloths spread on the ground or (more frequently today) on sheets of plastic.

Olives to be eaten are usually picked green and unripe. But, as Harold McGee (1984: 204) has remarked, anyone who has ever bitten into one knows instantly that something more must be done to render it edible. This something more is generally pickling, which has been practiced since the days of ancient Rome by soaking olives in a solution of lye to remove the bitter glucoside called oleuropein (from *Olea europea*).

Black olives – those that ripen at the time of the first frosts – are the kind made into oil. After they are gathered, they are permitted to stand and get warm, but not to ferment. Then they are washed and crushed. The oldest known technique for extracting oil was that of crushing the fruit underfoot; later, crushing was done by hand, using a pestle, and later still with millstones. In fact, many of the old-style Roman oil mills, with their large millstones operated by donkeys or mules (or by slaves in Roman times) continued in use throughout the Mediterranean area until well into the twentieth century (Toussaint-Samat 1992).

The crushing results in a paste that is pressed to secure the 25 to 30 percent of the olive that is oil. Extra virgin oil, also called “cold-pressed,” comes from the first and lightest pressing and is unrefined; moreover, no heat is used to extract further oil. Virgin oil is produced in the same manner but has slightly greater acidity. Cold-pressed olive oil, although it has a wonderful flavor, will not keep as well as refined olive oil and, consequently, must be shielded from the light in cans or dark bottles.

The oil produced by succeeding pressings (usually with heat or hot water added) generally contains substances that give it a bad flavor, whereupon it is refined. Light and extra light oils are olive oils (not virgin) that have been filtered, producing a light scent, color, and taste. Oil labeled “pure” is a combination of virgin oil and that derived from the second pressing. Production of olive pomace oil – the cheapest grade of olive oil – requires hot water, which is added to the residue of olive oil cake, then poured off. This oil, typically not very good, has traditionally been exported to countries where a taste for olive oil has not been highly developed, although it is sometimes consumed by its own producers, who in turn sell their virgin oils for income (Toussaint-Samat 1992).

Table II.E.1.2. *Fatty acid composition ranges (weight percentage) of natural populations of vegetable oils*

	Soybean	Palm	Sunflower	Rapeseed	Peanut	Cottonseed	Coconut	Olive
6:0	0	0	0	0	0	0	<1.2	0
8:0	0	0	0	0	0	0	3.4-15	0
10:0	0	0	0	0	0	0	3.2-15	0
12:0	0	<1.2	0	0	0	0	41-56	0
14:0	<0.5	0.3-5.9	<0.5	<1	<1	0.5-2	13-23	<0.5
16:0	7-12	27-59	4-9	0.5-10	6-14	17-29	4.2-12	7.5-20
16:1	<0.5	<0.6	<0.5	<1	<1	<1.5	0	0.3-3.5
18:0	2-5.5	1.5-14.7	1-6.5	0.5-4	2-6.5	1-4	1-4.7	0.5-3.5
18:1	20-50	27-52	14-70	9-58	40-72	13-44	3.4-12	56-86
18:2 ω 6	35-60	5-16	20-75	8-27	13-38	33-58	0.9-3.7	3.5-20
18:2 ω 3	2-13	0.5-5	<0.5	3-21	<0.5	<0.5	0	01.5
20:0	<1	<1	<1	<1.5	1-2	<0.5	<0.2	<0.1
20:1	<1	0	<0.5	5-18	0.5-1.5	<0.5	<0.2	<0.1
22:0	<0.5	0	<1	<1.5	2-4	0	0	<0.1
22:1	0	0	0	30-60	0	0	0	0
24:0	0	0	<0.2	<2	1-2	0	0	<0.1

A more complicated (and technical) grading system for olive oils is based primarily on the free fatty acid levels: Extra virgin oil, which comes from the first pressing, has an acidity of less than 1 percent and perfect flavor and odor. Virgin oil, from the same pressing, contains less than 3 percent free fatty acids. Extra virgin and virgin olive oils may be processed only by washing, decantation centrifugation, and filtration. Refined olive oils are obtained by refining virgin or extra virgin olive oils, using the same alkali neutralization, bleaching, and deodorization steps employed in refining other oilseeds. Olive pomace oil is produced by solvent extraction from the cake remaining after pressing. It was sometimes called sulfur olive oil in the past, when carbon disulfide was used as an extraction solvent.

To maintain low free fatty acid levels and produce oil of the highest quality, olives must be processed within three days of harvest. This is primarily because ripe olives contain a high amount of water, allowing lipase enzyme activity and a resulting increase in free fatty acid levels. Consequently, in some areas that are not equipped to press olives as fast as they are harvested, either the fruit is left on the trees until it can be processed, or it is stored, which can be a risky procedure.

Olive oil is characterized by a very high oleic acid content averaging around 75 percent (Table II.E.1.2). Palmitic acid is the second most prevalent fatty acid, followed by linoleic and linolenic. There has been a great deal of interest in olive oil as a promoter of human health since the development of the hypothesis that a "Mediterranean diet" is protective against heart disease, perhaps because for centuries, olive oil has been the only oil readily obtainable for human use that has a high oleic characteristic.

Recent developments in oilseed breeding and genetics, however, have produced high-oleic safflower, sunflower, peanut, canola, and soybean oils. Yet, it is unlikely that these will take over the applications of olive oil, which is enjoyed as much for its flavor as for any nutritional advantages it may have. As already noted, because they are not refined, virgin olive oils have a pleasant and characteristic flavor, and of all the major vegetable oils, only olive oil is sold unrefined. The deodorization process undergone by other oils removes compounds that would otherwise produce flavors unsuitable for many applications. In addition, some oils have flavors that can only be characterized as unpleasant.

Olive oil has a moderate amount of tocopherols, primarily alpha-tocopherol. Despite a low level of gamma-tocopherol, olive oil is remarkably stable, which is attributable to its low levels of linoleic and linolenic acids, as well as phenolic compounds that are present. Minor components in olive oil are not removed because the oil is not refined; the levels of some of these compounds are shown in Table II.E.1.3. Finally, it should be noted that olive oil contains very low levels of phospholipids and waxes, which permits the production of a clear oil without refining.

Table II.E.1.3. *Levels of minor components in olive oil*

Component	Mg percentage
Squalene	125-750
Sterols	125-250
Triterpene alcohols	50
Chlorophyll	0.06-0.22
Carotenoids	0.06-0.95

Soybean Oil

The soybean (*Glycine max* L., formerly *Soja max*), of the family Leguminosae, is grown worldwide for its high-protein content and its oil. There are wild (*Glycine ussuriensis*) and intermediate (*Glycine gracilis*) as well as cultivated (*G. max*) soybean varieties. Although soybeans have been used as food in China for thousands of years, their development as the world's primary oilseed has only taken place since the mid-1940s.

As mentioned (and as shown in Figure II.E.1.2), the production of soybean oil today overshadows all others. The soybean contains 40 to 50 percent protein and an average of 18 to 20 percent lipids (Table II.E.1.1). Its oil is a byproduct of the production of soybean meal, which is primarily used as animal feed. The oil obtained directly from soybeans contains about 88 percent neutral lipid (triacylglycerols and sterols), 10 percent phospholipid, and 2 percent glycolipid. Refined oil contains more than 99 percent triacylglycerols, and it is necessary to remove the phospholipids to produce a stable, clear oil. Like other vegetable oils, save that from olives, soybean oil is usually refined, bleached, and deodorized before use.

Soybeans contain a number of antinutritional compounds, including protease inhibitors, hemagglutinins, saponins, goitrogens, allergens, phytic acid, and oligosaccharides (raffinose, stachyose), as well as isoflavones (phytoestrogens). However, because these compounds are not transferred to the lipid, the oil is essentially devoid of them.

Soybean oil is one of the two common oils that contain appreciable levels of omega-3 fatty acids, and it is ironic that at the same time that some investigators were discovering the important roles played by omega-3 fatty acids in the human diet, other researchers were working to reduce the levels of these fatty acids in soybean oil (Hammond 1992). The reason for the interest in lowering linolenic acid in soybean oil (which ranges from 5.5 to 10 percent) is that it oxidizes at a rate some 2 or more times greater than linoleic acid, which, in turn, oxidizes at a rate about 10 times lower than oleic acid (Hammond 1994). Lowering linolenic acid levels results in a more

stable oil; the more unsaturated the oil, the less stable it will be. In the past, soybean oils have been lightly hydrogenated (brush hydrogenation) to reduce the linolenic acid content to less than 3 percent. This is uncommon today because of health interests in oils with no trans acids and also because of problems with cloudiness at refrigeration temperatures.

Recently, there has also been a great deal of interest in reducing the amount of palmitic acid in soybean oil, which is a major dietary source of it. This is because palmitic acid is thought to be one of the three saturated fatty acids most responsible for raising plasma cholesterol in humans (the other two are myristic acid and lauric acid). Varieties of soybeans with modified fatty acid composition have been developed, including low-linolenic, high-stearic, high-oleic, high-palmitic/low-linolenic, high-palmitic, and low-saturate/low-linolenic.

The low-saturate/low-linolenate composition, shown in Table II.E.1.4, has much lower palmitic acid levels than traditional soybean varieties; the other fatty acids are relatively constant. The linolenic and oleic acids are at the high end of the fatty acid ranges in normal soybeans. Lowering the palmitic acid content of soybeans has been a major improvement in terms of nutritional value, and more varieties of oilseeds with tailor-made fatty acid compositions for specific purposes should be seen in the future.

The tocopherol composition of soybean oil is shown in Table II.E.1.5. It should be noted that, in some reported analyses of vitamin E compounds, the beta and gamma isomers are combined, whereas other analyses do not report these isomers. The beta and gamma tocopherols are very similar in structure (Figure II.E.1.4). For most vegetable oils, the alpha and gamma isomers predominate. The alpha isomer is by far the most bioactive of the tocopherol isomers (Table II.E.1.6), but the antioxidant effect appears to be greater for the gamma isomer. Tocopherols are reduced during the refining process at the deodorization stage. Roughly 30 percent of the total tocopherols are removed. The data presented in Table II.E.1.6 are for refined, bleached, and deodorized (RBD) oils.

Table II.E.1.4. Fatty acid compositions (weight percentage) of modified fatty acid vegetable oils

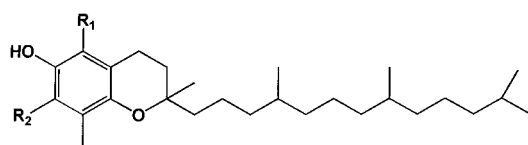
	High oleic peanut	Low erucic acid rapeseed (canola)	High oleic sunflower	Low saturate, low linolenate soybean
16:0	5.1	3.9	3.4	3.4
16:1	<0.1	0.2	0	0
18:0	2.3	1.3	4.8	3.4
18:1	80.4	58.2	81.0	32.5
18:2ω6	2.4	21.6	8.1	57.4
18:3ω3	<0.1	12.1	0.49	3.3
20:0	1.3	0.5	0.23	0
20:1	2.1	1.6	0.27	0
22:0	3.0	0.4	1.1	0
22:1	<0.1	<0.1	0	0
24:0	2.1	0	0.36	0
26:0	0.47	0	0	0

Table II.E.1.5. *Tocopherol isomer distribution in dietary fats and oils (μg/g)*

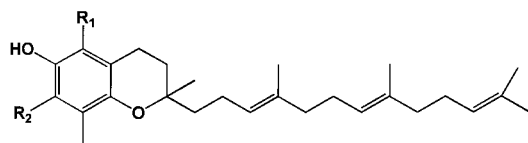
	α	β	γ	δ	α-tocotrienol
Soybean	75-116	34	662-797	266-400	2
Palm	288-360	-	280-360	80	146
Sunflower	110-610	10	30	10	-
Rapeseed	70-190	16	178-430	7.4-40	0.4
Peanut	60-169	5.4	100-200	13	-
Cottonseed	320	-	313	-	5
Coconut	20	-	2.4	-	-
Olive	70-154	-	7.3-25	1-3	-
Cod Liver	220	-	-	-	-
Herring	92	-	-	-	-
Menhaden	75	-	-	-	-
Lard	12	-	-	-	-
Tallow	27	-	-	-	-

Table II.E.1.6. *Approximate biological activity relationships of vitamin E compounds*

Compound	Percentage activity of <i>d</i> -α-tocopherol
<i>d</i> -α-tocopherol	100
<i>l</i> -α-tocopherol	26
<i>dl</i> -α-tocopherol	74
<i>dl</i> -α-tocopheryl acetate	68
<i>d</i> -β-tocopherol	50
<i>d</i> -γ-tocopherol	10
<i>d</i> -δ-tocopherol	3
<i>d</i> -β-tocotrienol	30



Tocopherol	R ₁	R ₂
alpha	CH ₃	CH ₃
beta	CH ₃	H
gamma	H	CH ₃
delta	H	H



Tocotrienol	R ₁	R ₂
alpha	CH ₃	CH ₃
beta	CH ₃	H
gamma	H	CH ₃
delta	H	H

Figure II.E.1.4. Tocopherol composition of soybean oil.

Palm Oil

Although there are many different palms that can provide oil, palm oil of commercial importance is obtained from the African oil palm, *Elaeis guineensis*. The name *Elaeis* is derived from the Greek word for oil. Palm oil is obtained from the fruit, whereas palm kernel oil comes from the seed kernel. Their fatty acid compositions are different, with palm kernel oil having a high content of lauric acid (palm kernel oil is similar to coconut oil in fatty acid composition). Palm oil is currently one of the most common oils worldwide, and oil amounts have increased dramatically since the 1980s.

Palm oil use in human diets can be dated back as far as 3000 B.C. Crude palm oil has a long history of use in western Africa, and in the eighteenth century, trade in the oil began between Africa and Europe. Wild palms are still a significant source of oil in West Africa, but recent increases in oil-palm cultivation have occurred in Latin America and Southeast Asia, with about 70 percent of present-day world palm oil production centered in the latter region.

Palm is the highest yielding of the oil-bearing plants, with an average of between 4 and 10 tons of oil per hectare per annum. Three types of palm have been characterized: *Dura*, with a thin flesh and thick, hard shell; *pisifera*, with thick flesh and little or no shell; and *tenera*, with a thick flesh and intermediate shell. The *tenera* variety is a cross between a *dura* female and pollen-bearing *pisifera*. Because the progeny of *tenera* are not uniform, cloning (asexual reproduction) of this palm has been practiced to improve yield and quality.

Traditional processing of palm to obtain palm oil, as practiced by the local populations, includes separation of the fruit, softening of the fruit, maceration, pressing, and oil purification. The softening process usually takes place by allowing the fruit to ferment in a pile for several days. The fruit may be further softened by boiling and then maceration in a mortar. After placement in a large oil-impermeable container, the oil is separated by kneading the material. Oil is skimmed from the top and may be filtered and then heated to remove the water.

The quantity of recovered oil in older traditional processing was low (less than 40 to 50 percent of total oil), and the quality of this crude oil was poor, compared to oil extracted and refined with more modern methods. If the fruits are not quickly heat treated, the free fatty acid level increases rapidly because of fungal lipases, and indeed, free fatty acid contents as high as 50 percent have been measured in traditionally processed oils. The flavor of oil with a high free fatty acid content is preferred in some areas of West Africa, where palm oils have traditionally been unrefined and have always had a high free fatty acid content. However, for most people, very high free fatty acid levels usually make the oil unsuitable for food use and more amenable to industrial applications, such as soap making. Oils obtained from wild trees by traditional processes are classified as soft (less than 12 percent),

Table II.E.1.7. *Nontriacylglycerol materials in crude palm oil*

Component	Concentration (ppm)
Carotenoids	500-700
Vitamin E compounds	699-1,000
Sterols and sterol esters	360-600
Squalane and aliphatic hydrocarbons	250-550

semisoft (less than 35 percent), or hard (greater than 45 percent), depending on the free fatty acid levels.

Crude palm oil contains a number of nontriacylglycerol components, including carotenoids, tocopherols and tocotrienols, sterols and sterol esters, phospholipids, and hydrocarbons (including squalane). The levels of these compounds in crude palm oil are shown in Table II.E.1.7. Palm oil contains very high levels of tocopherols and alpha tocotrienol (Table II.E.1.5). The carotenoids include lycopene and alpha and beta-carotene and are the cause of the deep red color of the oil. The main carotenoids are alpha- and beta-carotene, which make up over 90 percent of the total. Although there is experimental evidence linking dietary intake of carotene with protection against some types of cancer, the current practice is to remove the carotenes by bleaching to produce light-colored oil. Moreover, removing the carotenoids substantially improves the oxidative stability.

Recently, there has been a campaign virtually condemning the so-called tropical oils (palm, palm kernel, and coconut oils) as bad for human health. The argument has been that because they are high in saturated fats, these oils raise cholesterol levels in the body. Food products have been labeled "we use no tropical oils," their manufacturers and processors employing partially hydrogenated oils instead. Yet the notion that the addition of tropical oils to food products could lead to an increase in cardiovascular deaths ignores the fact that at least in Western diets (and especially in the United States), animal products and soybean oil are the main sources of saturated fats in the diet. It also ignores the cloud of suspicion now enveloping the use of partially hydrogenated oils.

The fatty acid composition of palm oil is quite different from that of coconut or palm kernel oil. The latter are lauric oils with about 80 to 90 percent saturated fat, predominantly lauric acid (41 to 55 percent), myristic acid (13 to 23 percent), and palmitic acid (4 to 12 percent). Palm oil is rich in oleic acid and low in saturates (less than 50 percent), relative to coconut and palm kernel oils. Experimental evidence so far indicates that there is no nutritional danger posed by the inclusion of palm oil in a Western diet, nor, for that matter, is such a danger posed by diets that derive a high percentage of total fat from palm oil (Berger 1994; Ong, Choo, and Ooi 1994). In fact, when palm oil is added to a Western diet, the level of plasma HDL cholesterol typically rises, leading to a

better LDL:HDL ratio, and this ratio – rather than the amount of total plasma cholesterol – appears to be the better indicator of the risk of coronary artery disease.

Palm oil can be fractionated to produce olein and stearin fractions; when heated to about 28° C, some 25 percent of the total remains solid (palm stearin). The liquid fraction (palm olein) has a greater unsaturated fatty acid content and can be further fractionated at 20-22° C (facilitated by the addition of a solvent) to yield a more liquid fraction, as well as an intermediate one. The fatty acid compositions of palm olein and palm stearin are shown in Table II.E.1.8. These fractions can be used for particular applications; for example, palm olein is commonly added to infant formulas because of its high oleic acid content. Plastic fats containing zero trans fatty acids are developed by interesterifying palm stearin and a more saturated fat, such as palm kernel oil. Palm olein has been used as a frying medium and has excellent stability because of its low linoleic and linolenic acid content and high levels of tocopherols. Cocoa butter equivalents can also be developed by using intermediate palm oil fractions.

Sunflower Oil

The sunflower (*Helianthus annuus*) is a wildflower that originated in the Americas. It was taken to Spain in 1569 for ornamental use, but later was grown for its oilseeds, especially in Russia. Newer, high-oil strains of sunflower have been developed, with oil contents upwards of 40 percent, as compared with 20 to 32 percent for traditional varieties.

Sunflower oil has a naturally high level of linoleic acid and is low in linolenic acid. The fatty acid composition of sunflower oil is highly variable and depends on climate, temperature, genetic composition, stage of maturity, and even the positions of individual seeds on the head of the flower. As the temperature of a region of sunflower cultivation increases, the oleic acid content increases and that of linoleic acid decreases, although together they always comprise about 90 percent of the fatty acids in sunflower oil.

Table II.E.1.8. *Fatty acid compositions (weight percentage) in palm olein and stearin*

Fatty acid	Olein	Stearin
12:0	0.1-1.1	0.1-0.6
14:0	0.9-1.4	1.1-1.9
16:0	37.9-41.7	7.2-73.8
16:1ω7	0.1-0.4	0.05-0.2
18:0	4.0-4.8	4.4-5.6
18:1ω9	40.7-43.9	15.9-37.0
18:2ω6	10.4-13.4	3.2-9.8
18:3ω3	0.1-0.6	0.1-0.6
20:0	0.2-0.5	0.1-0.6

A typical fatty acid composition in cold environments includes 14 percent oleic acid and 75 percent linoleic acid, whereas in warm environments, it is 50 percent oleic and 43 percent linoleic acid. Because most sunflowers are grown in moderate climates, however, the actual variation in sunflower oil composition is, as a rule, not as marked as it might be.

New varieties of sunflower oil have been developed with modified fatty acid compositions. The fatty acid composition of a sunflower oil high in oleic acid is shown in Table II.E.1.4. High-oleic sunflower oils originated in Russia in the mid-1970s as a result of selective breeding and mutagenesis. The original Russian high-oleic line, "Prevenets," has been developed into sunflower lines that have a stable high-oleic acid composition minimally affected by growing temperature (Morrison, Hamilton, and Kalu 1994). The high-oleic acid and low-linoleic acid composition makes the oil very suitable for frying, and the oxidative stability of products fried in high-oleic sunflower oil is excellent. However, flavor quality has been reported to suffer if the linoleic acid level is too low.

Crude sunflower oil contains a number of non-triglyceride components, including waxes, sterols, hydrocarbons, and tocopherols. Waxes, which create problems because they produce clouding of the oil, are removed in a process whereby the oil is cooled and the solids that develop are filtered off. The tocopherols in sunflower oil consist primarily of alpha-tocopherol, which is good from the viewpoint of human nutrition because this tocopherol is the most biopotent form of vitamin E. However, tocopherols stabilize oils from oxidation, and the gamma and delta forms are much more active antioxidants in vitro than the alpha. In sunflower oil, the low levels of these tocopherols produce a lower-than-expected oxidative stability when compared to other vegetable oils with a higher degree of unsaturation.

Rapeseed (Canola) Oil

Rapeseeds include a number of closely related species in the genus *Brassica*. The three main species (*Brassica nigra*, *B. oleracea*, and *B. campestris*) have been manipulated by hybridization and chromosome doubling to produce *B. napus*, *B. carinata*, and *B. juncea*. The different Brassica species have different seed sizes, colors, oil contents, and proximate compositions. The term "rapeseed" is used for any oil-bearing Brassica seed, including that of mustard. Common names for some rapeseed species are shown in Table II.E.1.9.

The use of rapeseed oil for food and illumination originated in Asia (later spreading to the Mediterranean and Europe), and today, it is the most widely used cooking oil in Canada, Japan, and some other countries. Cultivated in India at least 3,000 years ago, rapeseed was introduced into China and Japan by the time of Christ. By the thirteenth century, rapeseed was cultivated for oil in Europe – especially eastern

Table II.E.1.9. Common names for rapeseed species

<i>B. napus</i>	Rape, rapeseed, oil rape, oilseed rape, swede rape
<i>B. campestris</i>	Turnip rape, rapeseed, oil turnip, polish rape
<i>B. juncea</i>	Leaf mustard, Oriental mustard, Indian mustard, rapeseed
<i>B. nigra</i>	Black mustard
<i>B. oleracea</i>	Kale, cabbage

Europe – and recently, it has been grown in Canada and the United States. Its first use was probably as an oil for lamps and lubrication; in addition, the nourishing oil cake residue of rapeseed makes a good feed for cattle.

Research in the 1960s showed that the erucic acid found in rapeseed oil caused fatty infiltration of the heart and changes in the cardiac muscle in experimental animals. The oil was therefore banned in many places, and although subsequent work showed that rapeseed oil presented no threat to human health, there were intensive efforts to develop low-erucic acid cultivars of the plant. In 1985, the U.S. Food and Drug Administration (FDA) conferred a GRAS ("Generally Recognized As Safe") status on rapeseed products; not only was the oil deemed safe, but because of its very low content of saturated fat, it suddenly became viewed as the most healthy of cooking oils. Following the GRAS designation, U.S. production increased greatly, from 27 million pounds in 1987 to almost 420 million by 1994 (Trager 1995). In addition to its own production, the United States imports huge amounts of the oilseeds from Canada, although the end product's name was changed, understandably, from rapeseed to Canola oil.

Recent developments in rapeseed breeding have resulted in seeds that are modified in fatty acid composition, and low-erucic acid rapeseed (LEAR) oils have been produced by Canadian breeding work with *B. napus* and *B. campestris* (Table II.E.1.4). Double-low oils also are low in glucosinolates and include Canola (formerly Canbra) oil. To be considered a Canola or LEAR oil, the erucic acid content must be lower than 3.5 percent, although typically the level is less than 1 percent.

Canola oils are similar to soybean oils in that they contain appreciable levels of linolenic acid. Although some other vegetable oils have higher levels (for example, oils from flax seed or linseed contain about 55 percent linolenic acid), these are not common in the diet and are subject to very rapid rates of oxidation. In no small part because it has the lowest saturated fat content of any major cooking oil, Canola has been called an ideal mixture for health, nutrition, and food use (Ackman 1990). The oil has high amounts of oleic acid, reasonably low linoleic acid, and a moderate omega-6 to omega-3 ratio. The high-oleic level is

beneficial nutritionally, as oleic acid tends to lower LDL cholesterol, with no drop – and even a slight increase – in HDL cholesterol, resulting in a net positive change in the HDL:LDL ratio.

Peanut Oil

The peanut or groundnut (*Arachis hypogea*), like the bean, is a member of the family Leguminosae. It is native to South America, probably originating in or around lowland Bolivia, but as early as 3100 to 2500 B.C., peanuts were common on the coast of Peru, and by the time of the Columbian voyages, they had spread throughout much of the New World (Coe 1994). In the sixteenth century, the peanut was carried to West Africa from Brazil by Portuguese traders, and because it was first brought to the United States from West Africa in slave ships, the misconception arose that the plant was of African origin.

While the Portuguese moved the peanut eastward, the Spaniards carried it west. It was introduced into the Philippines by Ferdinand Magellan and spread thereafter to other parts of Asia. Peanuts can be grouped into four different types: Runner, Spanish, Virginia, and Valencia. Runner peanuts are small and high yielding, Spanish are small seeded, and Virginia are large seeded. Valencia are large seeded and are often found with 3 to 5 seeds per pod.

Normal peanut oil has almost no linolenic acid and about 25 percent linoleic acid. As a rule, it contains about 50 percent oleic acid and is considered a premium oil for food use because of its excellent stability and its pleasant, nutty flavor. As with some other oilseeds, temperature affects the fatty acid composition, with cool climates resulting in greater linoleic and lower oleic acid levels. High-oleic peanuts were developed at the University of Florida during the 1980s and have subsequently become commercially available. A typical fatty acid composition of a high-oleic peanut oil is shown in Table II.E.1.4. Peanut oil, including the high-oleic variety, is characterized by relatively high levels of long-chain saturated fatty acids (up to 6 percent), up to and including 26:0. High-oleic peanut oil is between 7 and 10 times more stable in the face of oxidation than normal peanut oil (O'Keefe, Wright, and Knauft 1993), and roasted high-oleic peanuts have about twice the shelf life of normal peanuts (Mugendi, Braddock, and O'Keefe 1997; Mugendi et al. 1998).

P. J. White (1992) has pointed out that in the past, peanut oil was wrongly reported to contain significant levels of linolenic acid because of the mistaken identification of 20:0 or 20:1ω9 as linolenic acid. Depending on analytical conditions, it is possible to have coelution of linolenic and 20:0 or 20:1ω9. Similarly, reports exist in which arachidonic acid (20:4ω6) has been mistakenly identified in vegetable oils, but, in fact, arachidonic acid has never been found in vegetable oils.

Cottonseed Oil

Cotton (any of the various plants of the genus *Gossypium*) has been grown for its fiber – used in the manufacture of rope and textiles – for close to 5,000 years. Doubtless, in the distant past, the oil-rich cotton seeds were also utilized for illumination, as a lubricant, and probably for medicinal purposes. Indeed, most of the vegetable oils seem to have been first utilized for such nonfood applications.

The widespread use of cottonseed oil, however, is a relatively recent phenomenon that began in the United States. Some of the oil is derived from the extra-long staple *G. barbadense*, but the overwhelming bulk of the world's production (90 to 98 percent) comes from the short staple *G. hirsutum*, and short-staple cotton had to await the invention of the cotton gin by Eli Whitney in 1793 to become commercially viable. Thereafter, short-staple cotton spread across the southern United States, and surplus seed was converted into oil. Initial attempts at commercialization, however, were unsuccessful: Unlike long-staple cotton, the short-staple species had a tough seed coat that retained lint.

This problem was resolved in the middle 1850s with the invention of a machine to crack the hulls. High oil prices encouraged cottonseed oil extraction, and in the late 1880s, the Southern Cotton Oil Company was founded in Philadelphia. Its many seed-crushing mills, located throughout the South from the Carolinas to Texas, made it the largest producer of cottonseed oil, as well as the first U.S. manufacturer to make vegetable shortening.

One problem that remained was the odor of cottonseed oil, but this was eliminated at about the turn of the twentieth century with a deodorizing process invented by the company's chemist David Wesson. The procedure involved vacuum and high-temperature processing and led to the production of Wesson Oil, which revolutionized the cooking-oil industry (Trager 1995). Thereafter, cottonseed oil became the first vegetable oil to be used in the United States and then the dominant oil in the world's vegetable oil market until the 1940s.

Cottonseeds contain gossypol, which is produced in the seed glands and must be removed or destroyed in both oil and meal. Gossypol is heavily pigmented and can produce a dark color in the oil, and it is also toxic. Fortunately, it can be almost completely removed during the refining process, and in addition, glandless seeds devoid of gossypol have been developed.

Crude cottonseed oil contains as much as 1 percent of cyclic fatty acids. The main fatty acids are malvalic and sterculic, which are present in the form of glycerides. Cottonseed oil is characterized by a high saturated fat level, which is much higher than that of corn, canola, soybean, sunflower, peanut, and safflower oils.

Coconut Oil

The coconut palm (*Cocos nucifera*) has been enthusiastically described as the “tree of life” or the “tree of heaven” because it provides leaves and fiber for the construction of dwellings and shells for utensils, as well as food, milk, and oil. Although the coconut palm has been exploited by humans for millennia, its origin is subject to debate, with some scholars claiming it to be native to South America and others supporting a Southeast Asian origin. The question is a difficult one because coconuts can still germinate after floating on the ocean for months, a capability which has facilitated their spread throughout the tropical world. Compared with other oilseed crops, coconuts have a high per-acre yield of oil, about half of which is used in nonfood applications (such as soaps and cosmetics), whereas the other half goes into foods.

Coconut oil is extracted from the dried meat of the nuts (copra). Copra has one of the highest fat contents of all the materials used for oil (Table II.E.1.1), making it easy to obtain oil by pressing. Coconut oil contains high levels of the medium chain saturated fatty acids (10:0, 12:0, and 14:0) and is solid at room temperature.

The saturation of coconut oil confers a high degree of stability toward oxidation, but hydrolysis of coconut oil to yield free fatty acids causes soapy off-flavors because of the free lauric acid present. Fatty acids longer than lauric acid, however, produce little or no off-flavor.

Animal Fats and Oils

Milkfat (Butter)

Milkfats are derived from any number of different animals, but most by far is provided by the bovine; consequently, in this chapter, milkfat means that from cows. A lactating cow can produce as much as 30 kg of milk per day with a milkfat level of 2.6 to 6 percent, and thus for many years, butterfat was produced in greater quantity than any other fat or oil.

Milkfat, which has a much more complex composition than any of the other common fats, contains about 97 to 98 percent triglycerides, 0.1 to 0.44 percent free fatty acids, 0.22 to 0.41 percent cholesterol, and 0.2 to 1.0 percent phospholipids. More than 500 different fatty acids have been identified in milk lipids. The fatty acid patterns in milkfat are complicated by mammary synthesis of short-chain fatty acids and extensive involvement of rumen bacteria in lipid synthesis and modification (Jensen 1994). As a rule, butterfat is predominantly composed of palmitic acid, followed by oleic and myristic acids (Table II.E.1.10). About 66 percent of the fatty acids in a typical milkfat are saturated. The trans fatty acid levels in milkfat have reportedly ranged between 4 and 8 percent.

Table II.E.1.10. Fatty acid composition of butterfat (weight percentage)

Fatty acid	Range
4:0	2.8–4.0
6:0	1.4–3.0
8:0	0.5–1.7
10:0	1.7–3.2
12:0	2.2–4.5
14:0	5.4–14.6
16:0	26–41
18:0	6.1–11.2
10:1	0.1–0.3
12:1	0.1–0.6
14:1ω5	0.6–1.6
16:1ω7	2.8–5.7
18:1ω9	18.7–33.4
18:2ω6	0.8–3.0

There has been a great deal of interest over the past 30 years in modifying the fatty acid composition of butterfat to decrease the saturated fat content (Sonntag 1979b). Because cows have a rumen, fats that are directly provided in the diet are extensively degraded and hydrogenated. The fat also can have negative effects on the rumen bacteria, causing a lowering of milk and milkfat production. This problem has been solved by providing “protected” fats to the animal.

Methods of protecting fats from rumen metabolism include the polymerizing of protein around fat with formaldehyde and the preparation of calcium soaps, which are mostly insoluble at rumen pH. Dairy fats with high linoleic acid contents have been developed, but they suffer from severe quality problems and are unmarketable (McDonald and Scott 1977).

Recently, dairy fats have been prepared with high oleic and low saturate composition by feeding a calcium salt of high-oleic sunflower oil (Lin et al. 1996a, 1996b). This has provided lower saturated fatty acid composition and higher monoene while maintaining product quality. Oleic acid oxidizes about 10 times more slowly than linoleic acid and provides a better health benefit.

Marine Fats

Marine fats and oils, which are among the oldest exploited by humans for food and other uses, are divided into those from fish (cod, herring, menhaden, sardine, and so forth), from fish livers (cod, shark, halibut), and from marine mammals (whales). Marine oils are characterized by a very complex composition compared to vegetable oils and many animal fats, which poses challenges in the accurate identification and quantification of their fatty acids. Moreover, data derived from old technology may not be accurate.

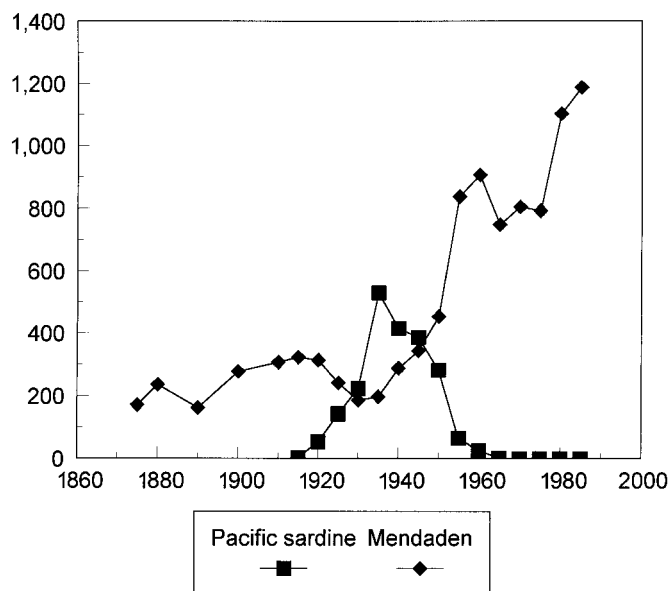


Figure II.E.1.5. Effects of overfishing in Pacific sardine fishery, 1918-60.

Fish oils. Fish oils are obtained from tissues that contain a high fat content, including those of herring, pilchard, menhaden, and anchovy. The production of the various fish oils is dependent on both fishing efforts and overfishing. In Figure II.E.1.5, it can be seen that the Pacific sardine fishery was extremely successful for about 20 years, after which the fish stocks were depleted so that a fishery could no longer be supported. It is sad to note that this is the rule rather than the exception in many fisheries worldwide.

Early production methods consisted of boiling the fish in water, followed by pressing, and then removing the oil by skimming. Improved oil quality and recovery was obtained as steam cooking and screw presses were introduced. Much of the fish oil produced in earlier days went into nonfood uses, such as linoleum, paints, and soaps.

Currently, oils from marine sources constitute some 2 to 4 percent of the world's consumption of fats, and about one-third of the world's fish catch consists of small and bony but high-fat fish that are converted into fish meal and oil. Fish oils for food use are often partially hydrogenated because they oxidize rapidly in comparison with vegetable or animal fats and oils. Partial hydrogenation results in a loss of polyunsaturated fatty acids, monoenes, dienes, and trienes, which obviously has negative effects on the nutritional value of hydrogenated oils. Because of the interest in omega-3 fatty acids found in fish oils, recent work has focused on the incorporation of processed, but not hydrogenated, fish oils in food products.

Fish liver oils. The livers of some fish have a high fat content and are rich in vitamins A and D. Although

many different species of fish (for example, pollock, haddock, shark, halibut) have been exploited for their liver oils, the cod has provided more than 90 percent of the total volume. Cod livers contain from 20 to 70 percent fat; the range is so great because codfish deposit fat prior to spawning and deplete it during spawning. Thus, there is a cyclical fat content in their livers, and this has a dilution effect on the levels of the fat-soluble vitamins.

Cod liver oil was reportedly used for medicinal purposes in the eighteenth century, long before the discovery of vitamins or an understanding of most nutritional diseases. Medical uses included treatment of bone afflictions, rheumatism, tuberculosis, and, later, rickets. There was an intensive search for the active ingredient in cod liver oil that caused such positive therapeutic effects. The hunt was described by E.P. Moller (1895: 69):

The oldest of these active principles, iodine, was detected as one of the oil's constituents as far back as 1836. . . . When the iodine active principle was exploded another, trimethylamine, took its place, and as a result herring brine, from being a rather neglected commodity, became for a short time the desideratum of the day. After a quick succession of other active principles, too evanescent to be worthy of even an obituary notice, the turn came for the free fatty acids. No theory and no fact then supported the idea that these acids were the active principle. . . . Still, the belief entered some active mind that they were the essential constituents.

Moller concluded incorrectly that there was no specific active principle in cod liver oil. Yet, liver oils

provide fat-soluble vitamins. However, when consumed at the levels required to derive significant health benefit from the fatty acids present in fish liver oils, there is a risk of hypervitaminosis A or D.

The levels of fat-soluble vitamins in various fish liver oils are shown in Table II.E.1.11. Excessive levels of both vitamins A and D₃ can be toxic. Although the actual value is not clear, as little as 50 micrograms of vitamin D₃ per day over extended periods appears to produce toxicity symptoms in humans (Groff, Gropper, and Hunt 1995). Daily levels of 250 micrograms for several months have resulted in soft-tissue calcification, hypertension, renal failure, and even death. The recommended daily intake of vitamin D is 5 to 10 micrograms. The vitamin D level in fish oils ranges up to 750 micrograms per gram. Obviously, extended daily use of a high-potency fish oil should be avoided. Cod liver oil, which constitutes the majority of all liver oils sold today, has modest but nutritionally significant levels ranging up to 7.5 micrograms per gram.

The recommended daily intake of vitamin A is 1,000 retinol equivalents. One retinol equivalent is equal to 1 microgram of all-trans retinol. Toxicity is seen at levels around 10 times greater than the RDA, so that a daily intake of 10 mg of all-trans retinol would result in toxic symptoms. Symptoms of toxicity include itchy skin, headache, anorexia, and alopecia. The extremely high vitamin A levels in some liver oils would make it fairly easy to consume toxic amounts of vitamin A.

Whale oils. Whale oils are derived from marine mammals. The worldwide production of whale oils has decreased as populations of whales have been depleted. Oils from the baleen whales (sei, right, blue, humpback, and minke), as well as toothed and sperm whales, have been employed in industrial applications and human food use. Sperm whales (*Physeter macrocephalus*), in particular, were prized in the past for their "spermaceti" oil, which was used in smokeless lamps, as a fine lubricant, in cosmetics, and in automatic transmission oils. One sperm whale could yield 3 to 4 tons of spermaceti.

Table II.E.1.11. *Fat-soluble vitamin levels in fish liver oils*

Fish liver oil	Vitamin A, retinol equivalents/g	Vitamin D ₃ µg/g
Cod	165-3,000	0.5-7.5
Haddock	45-900	1.3-1.9
Dogfish	60-45,000	0.1-0.7
Halibut	1,200-49,500	13.8-500
Mackerel	900-49,000	19-25
Tuna	15,000-300,000	400-750
Hake	480-960	0.25-3.3
Swordfish	6,000-120,000	50-625

Whale oils saw food use in some, but not all, countries when whale harvests were high. Europe, Canada, and Japan, for example, utilized hydrogenated whale oil in margarines for almost 50 years. During that time, the production of fish oils was much lower, increasing only as the production of whale oil decreased. Whale oils were also employed in the manufacture of shortenings after hydrogenation, a process which results in the production of many trans fatty acid monoene isomers.

Conclusion

The types and amounts of fats and oils that have been consumed vary greatly worldwide among people and have changed significantly over the past century. Fatty acid intakes will undoubtedly change in the future as oilseeds with altered fatty acid compositions reach new markets.

Sean Francis O'Keefe

Bibliography

- Ackman, R. G. 1989. Problems in fish oils and concentrates. In *Fats for the future*, ed. R. C. Cambie, 189-204. Chichester, England.
1990. Canola fatty acids - An ideal mixture for health, nutrition and food use. In *Canola and rapeseed*, ed. F. Shahidi, 81-98. New York.
1992. Fatty acids in fish and shellfish. In *Fatty acids in foods and their health implications*, ed. C. K. Chow, 95-135. New York.
- Ackman, R. G., S. N. Hooper, and D. L. Hooper. 1974. Linolenic acid artifacts from the deodorization of oils. *Journal of the American Oil Chemists' Society* 51: 42-9.
- Ackman, R. G. 1989. Marine mammals. In *Marine biogenic lipids, fats, and oils*, 2 vols. Vol. 2, ed. R. G. Ackman, 179-381. Boca Raton, Fla.
- Bailey, B. E. 1952. Marine oils with special reference to those of Canada. *Fisheries Research Board of Canada Bulletin* 89: 1-413.
- Berger, K. 1994. Oils from under-utilized palm and forest products. In *Technological advances in improved alternative sources of lipids*, ed. B. S. Kamel and Y. Kakuda, 209-34. New York.
- Bimbo, A. P. 1990. Production of fish oil. In *Fish oils in nutrition*, ed. M. E. Stansby, 141-80. New York.
- Braddock, J. C., C. A. Sims, and S. F. O'Keefe. 1995. Flavor and oxidative stability of roasted high oleic peanuts. *Journal of Food Science* 60: 489-93.
- Chandler, William H. 1950. *Evergreen orchards*. Philadelphia, Pa.
- Chen, I. C., C. I. Wei, F. A. Chapman, et al. 1995. Differentiation of cultured and wild Gulf of Mexico sturgeon (*Acipenser oxyrinchus desotoi*) based on fatty acid composition. *Journal of Food Science* 60: 631-5.
- Coe, Sophia D. 1994. *America's first cuisines*. Austin, Tex.
- Formo, M. W. 1979. Fats in the diet. In *Bailey's industrial oil and fat products*, Vol. 1., ed. D. Swern, 233-70. New York.
- Groff, J. L., S. S. Gropper, and S. M. Hunt. 1995. *Advanced nutrition and human metabolism*. St. Paul, Minn.

- Hammond, E. G. 1992. Genetic alteration of food fats and oils. In *Fatty acids in foods and their health implications*, ed. C. K. Chow, 313-27. New York.
1994. Edible oils from herbaceous crops. In *Technological advances in improved alternative sources of lipids*, ed. B. S. Kamel and Y. Kakuda, 93-115. New York.
- Holman, R. T. 1992. A long scaly tale - The study of essential fatty acid deficiency at the University of Minnesota. In *Essential fatty acids and eicosanoids*, ed. A. Sinclair and R. Gibson, 3-17. Champaign, Ill.
- Hunter, J. E. 1992. Safety and health effects of isomeric fatty acids. In *Fatty acids in foods and their health implications*, ed. C. K. Chow, 857-68. New York.
- Jensen, R. G. 1992. Fatty acids in milk and dairy products. In *Fatty acids in foods and their health implications*, ed. C. K. Chow, 95-135. New York.
- Jones, L. A., and C. C. King, eds. 1993. *Cottonseed oil*. National Cottonseed Products Association. Memphis, Tenn.
- Lin, M. P., C. A. Sims, C. R. Staples, and S. F. O'Keefe. 1996a. Flavor quality and texture of modified fatty acid high monoene low saturate butter. *Food Research International* 29: 367-72.
- Lin, M. P., C. Staples, C. A. Sims, and S. F. O'Keefe. 1996b. Modification of fatty acids in milk by feeding calcium-protected high oleic sunflower oil. *Journal of Food Science* 61: 24-7.
- McDonald, I. W., and T. W. Scott. 1977. Food of ruminant origin with elevated content of polyunsaturated fatty acids. *World Review of Nutrition and Dietetics* 26: 144-207.
- McGee, Harold. 1984. *On food and cooking*. New York.
- Moller, F. P. 1895. *Cod-liver oil and chemistry*. London.
- Morrison, W. H., R. J. Hamilton, and C. Kalu. 1994. Sunflower seed oil. In *Developments in oils and fats*, ed. R. J. Hamilton, 132-52. New York.
- Mugendi, J. B., J. C. Braddock, and S. F. O'Keefe. 1997. Flavor quality and stability of high oleic peanuts. In *Chemistry of novel foods*, ed. O. Mills, H. Okai, A. M. Spanier, and M. Tamura, 153-66. Carol Stream, Ill.
- Mugendi, J. B., C. A. Sims, D. W. Gorbet, and S. F. O'Keefe. 1998. Flavor stability of high oleic peanuts stored at low humidity. *Journal of the American Oil Chemists' Society* 74: 21-5.
- Nawar, W. W., S. K. Kim, Y. J. Li, and M. Vajdi. 1991. Measurement of oxidative interactions of cholesterol. *Journal of the American Oil Chemists' Society* 68: 496-8.
- O'Keefe, S. F., D. Wright, and D. A. Knauff. 1993. Comparison of oxidative stability of high- and normal-oleic peanut oils. *Journal of the American Oil Chemists' Society* 70: 489-92.
- O'Keefe, S. F., D. Wright, and V. Wiley. 1993. Effect of temperature on linolenic acid loss and 18:3 cct formation in soybean oil. *Journal of the American Oil Chemists' Society* 70: 915-17.
- Ong, A. S. H., Y. M. Choo, and C. K. Ooi. 1994. Developments in palm oil. In *Developments in oils and fats*, ed. R. J. Hamilton, 153-91. New York.
- Salunkhe, D. K., J. K. Chavan, R. N. Adsule, and S. S. Kadam. 1992. *World oilseeds, chemistry, technology and utilization*. New York.
- Smith, L. L. 1987. Cholesterol oxidation. *Chemistry and Physics of Lipids* 44: 87-125.
- Sonntag, N. O. V. 1979a. Sources, utilization and classification of oils and fats. In *Bailey's industrial oil and fat products*, Vol. 1, ed. D. Swern, 271-88. New York.
- 1979b. Composition and characteristics of individual fats and oils. In *Bailey's industrial oil and fat products*, Vol. 1, ed. D. Swern, 289-478. New York.
- Stansby, M. E. 1978. Development of fish oil industry in the United States. *Journal of the American Oil Chemists' Society* 55: 238-43.
- Tannahill, Reay. 1988. *Food in history*. New York.
- Thomas, L. M., and B. J. Holub. 1992. Nutritional aspects of fats and oils. In *Fatty acids in foods and their health implications*, ed. C. K. Chow, 16-49. New York.
- Toussaint-Samat, Maguelonne. 1992. *A history of food*, trans. Anthea Bell. Cambridge, Mass.
- Trager, James. 1995. *The food chronology*. New York.
- Ward, Artemas. 1911. *The grocer's encyclopedia*. New York.
- White, P. J. 1992. Fatty acids in oilseeds (vegetable oils). In *Fatty acids in foods and their health implications*, ed. C. K. Chow, 237-62. New York.

II.E.2 🌴 Coconut

Milk Bottle on the Doorstep of Mankind

In prehistoric times, the water content of the immature coconut fruit was more important as a drink than was any part of the mature nut as a food. In recent history, the emphasis has also been on a nonfood use of coconuts as oil. The oil extracted from the kernel of the ripe coconut is an industrial raw material for products ranging from soap to explosives. From prehistory to the present, coconut has served many human communities around the tropics in a variety of ways. In 1501, King Manuel of Portugal itemized some of its uses at a time when the coconut was first becoming known in Europe: "[F]rom these trees and their fruit are made the following things: sugar, honey, oil, wine, vinegar, charcoal and cordage . . . and matting and it serves them for everything they need. And the aforesaid fruit, in addition to what is thus made of it, is their chief food, particularly at sea" (Harries 1978: 277).

Unfortunately, it is not possible to provide as much information as one might want on the coconut in prehistory. This is because heat and humidity work against the preservation of fossils, and thus there is a dearth of archaeological materials, coprolites, and biological remains on tropical seashores where the coconut palm is native. Coconut residues do not accumulate because the palm grows and fruits the year round. This makes crop storage unnecessary and, in fact, because of their high water content, coconut seednuts cannot be stored; they either grow or rot. And the tender, or jelly, coconut is even less likely to survive in storage.

The sweet liquid in the immature fruit, however, is safe to drink where ground water may be saline or contaminated. It is a very pleasant drink, and coconuts are readily transported by land or sea.

In short, coconut is potable, palatable, and portable! Unlike drinks that are bottled, canned, or otherwise packaged, coconuts are sustainable and recyclable. It has been suggested that as the “milk bottle on the door step of mankind” (Harries 1981), the coconut could have played a significant role in the diets of our human ancestors in the time before agriculture.

Leafy Vegetables, Fruits, and Nuts

Millionaire's Salad

Although not strictly a green leafy vegetable, coconut heart of palm can be compared with blanched leafy vegetables, such as endive, or celery, or globe artichoke. It has been called “millionaire's salad” on the assumption that only the very rich can afford to fell an entire palm and have the leaf stalks cut away to expose the large bud, which is the part that is eaten. Palm hearts are best eaten fresh, but they can be cooked, canned, or pickled (Harries 1993).

Other palms can also be employed in this manner; indeed, with some palm species, harvesting heart of palm is a commercial operation. Certainly palm heart production could easily be commercialized for the coconut palm, especially in tropical coastal areas where tourism has replaced indigenous agriculture. One coconut heart may account for 40 side salads, and over-aged palms, overgrown seedlings from coconut nurseries, and those sprouting like weeds in neglected groves could be used for this purpose. It would even be practical to plant them at high density for sequential harvesting.

Farmers, however, are reluctant to cut down coconut palms, even when these are over-aged. They do not thin out palm stands that are too dense to be productive, and they usually ignore overgrown seedlings. All this may be attributable to a past in which the coconut palm was potentially the sole surviving food plant after a tidal wave or hurricane. Thus, the notion persists among some that to cut down a coconut palm threatens future life support. Moreover, in some communities, coconuts are planted to celebrate a birth, and if the palm dies or is felled, the human life it was planted to commemorate may be jeopardized. A recent example of the extreme reluctance to cut down the trees occurred during Liberia's civil war, when coconut palm hearts were eaten by the starving population only as a last resort.

Apple for the Teacher

Botanically, the coconut fruit is a drupe. Plums, peaches, and cherries, which are also drupes, have edible outer parts to encourage dissemination by animals. Other palms as well, particularly the date, have soft, sweet, and edible fruit, but the coconut is different because the outer covering, the husk, is generally bitter and stringy when young and dry and fibrous when mature. However, some rare individual coconut palms have an edible husk that is less fibrous,

spongier, easily cut, and sweet to chew like sugar cane (Harries 1993).

The coconut “apple” is, botanically, the haustorium of the germinating seed. The haustorium begins to develop at the earliest stage of germination, even before the shoot or roots emerge through the husk. Coconuts harvested in this condition are suitable for domestic purposes or for second-grade copra, but generally not for desiccated coconut. Often the apple is put aside to eat. It is slightly sweet, and slightly oily, with a cottony texture. As the endosperm lasts up to 15 months during germination, a large apple is found in well-developed seedlings. In places where coconuts grow, children walking to school may grasp the leaves of a sprouted seednut and uproot it. Still holding it by the leaves, they swing it against the trunk of the nearest mature palm to split the husk and crack open the shell. Then they pick out and eat the apple (Harries 1993).

An unusual form of the mature coconut has a jelly-like endosperm. This can be eaten fresh, scooped with a spoon from the shell of the freshly cracked coconut. It is called *makapuno* in the Philippines, where it is highly esteemed, and *dikiri-pol* or similar names in India and Sri Lanka. It is known in other coconut-growing countries, such as Indonesia, and has been reported in the Solomon Islands. The most interesting fact about it is that the embryo is normal but can only be germinated under the artificial conditions of a tissue culture laboratory.

A coconut with aromatic endosperm, favored in Thailand for drinking and preparing a cooked dessert, is also known in Malaysia and Indonesia.

Lovely Bunch of Coconuts

The coconut has at times been treated as something of a joke by Europeans, and it was popularized in the music hall song, “I've got a luvverly bunch of coconuts.” Historically, it was first introduced to Europe as *Nux indica*, or the nut from the Indies. This was possibly a generic name applied to the shells of all other palms that survived the long overland journey; and it may have referred to nutmegs as well (Ridley 1912). Such nuts and shells were kept as novelties, and even ornamented (Fritz 1983).

Friar John of Montecorvino (around 1292) described *Nux indica* as “big as melons, and in colour green, like gourds. Their leaves and branches are like those of the date palm” (quoted in Desmond 1992: 9). But it was not until after the Portuguese sailed to the Indian Ocean in 1499, and brought back fresh samples, that the coconut was distinguished from other palm fruits and from the nutmeg (although people in the countries where these plants grow cannot imagine that such a confusion could exist).

After the fibrous husk is removed, the brown, hard-shelled nut can be split to expose the kernel. Unlike an almond, for example, which also has a fibrous outer covering, a shell, and a kernel, the coconut is

generally not used as a nut because of its large size, although health food and vegetarian shops often include slices of coconut kernel in packets of mixed nuts. In England, coconuts sold in greengrocers' shops (without husks) are usually split open and hung outside for birds to feed on, especially in winter.

Oils, Fats, and Food

Fish and Chips

Coconut oil is most certainly a part of the diet in the countries where it grows. But equally importantly in those countries, it may be an unguent for the hair, an emollient for the body, a rust inhibitor for iron, and a fuel for lamps. Its first industrial use in Europe was as a lubricant in textile mills, although it subsequently became important to soap makers. Some of the latter are among today's industrial giants, and they still import coconut oil for the excellent lathering properties it imparts. It is interesting to note, however, that when soap manufacturers began using coconut oil, they unintentionally fostered the fish and chip shop.

This famous institution, the forerunner of all fast-food takeaways, became part of the social fabric in Britain (Harries 1988). Fish and chips date from the middle of the nineteenth century. Before then, local "soap-boilers" accumulated animal fat as the major ingredient of laundry soap. But animal fat was chiefly available in the winter months, when animals were slaughtered if they could not be fed, whereas coconut oil was available year-round from overseas colonial possessions that had an abundance both of the crop and of cheap labor.

Whether for soap or for cooking, coconut oil was particularly acceptable because it was convenient to handle. In a cool climate it does not even look like oil; below 20° to 26° C it becomes a greasy, somewhat crystalline, white or yellowish solid fat. In other words, outside the tropics, coconut oil becomes solid and resembles animal fat. It was also a good substitute for animal fat because there was no risk from infectious disease in its production.

The virtues of coconut oil were extolled at the beginning of the twentieth century in an advertisement for "Nut Lard" as:

an absolutely pure vegetable fat, extracted from the coco-nut. It is sweeter than ordinary lard or butter, and cheaper than either. It is white, odourless, does not turn rancid and is infinitely superior to ordinary lard for all culinary purposes. It can be used with the most delicate dishes without altering the natural flavour of the dish.

This advertisement went on to state that "'Nut Lard' contains neither salt nor water . . . [and] [i]n cold weather, 'Nut Lard' may become hard - it should then be shredded before using. . . ." The most telling part of the advertisement was that "'Nut Lard' is unequalled

for frying fish, it does not splutter, there is no smell, and it can afterwards be strained and kept for future use" (Anon. 1912: 41-2).

It was in the nineteenth century that such flamboyant advertising became a significant factor in marketing. In particular, industrial soap makers started large-scale advertising of their products, with the prepackaged brand names that still survive today. Such competition put the small, local soap boilers out of business, and when they could no longer sell soap, they looked for something else to do with their existing equipment - deep copper pans over open fires. The coconut oil they had previously used for soap was now put to work frying fish and chips.

Coconut Oil

In coconut-growing countries, coconut oil is prepared in the home by heating coconut milk until a clear oil separates. Commercially, the extraction of oil from copra (dried or smoked coconut meat) is one of the oldest seed-crushing industries in the world. Extraction methods range from simple techniques employed in villages to modern high-pressure expellers and prepress or solvent extraction plants that can process more than 500 tons of copra a day. In Indonesia, some processors cook chopped fresh kernel in previously extracted coconut oil before pressing. Various methods have been developed for "wet" processing of edible grade oil and flour from fresh meat, but none are yet commercially viable (Harries 1993).

Coconut oil is the most important of the small group of commercial fats that contain a high proportion of glycerides of lower fatty acids. The chief fatty acids are lauric (45 percent), myristic (18 percent), palmitic (9 percent), oleic (8 percent), caprylic (7 percent), capric (7 percent), and stearic (5 percent). There is also a minute amount of tocopherol (vitamin E). The natural volatile flavor components of fresh meat and oil are mostly delta-lactones. Lauric oils are characterized by high saponification value and have the lowest iodine value of vegetable oils in common industrial use. Coconut oil, as it is ordinarily prepared in tropical countries, ranges from colorless to pale brownish yellow in hue. In temperate climates, or air conditioning, it appears as a greasy, somewhat white or yellowish, solid fat that has a melting point between 20° and 26° C. Until refined, it has a pronounced odor of coconut. Coconut oil is refined, bleached, and deodorized using standard vegetable oil-processing technology. If coconut oil is cooled until crystallization, part of the oil produces a semi-solid mass, which is then separated under hydraulic pressure.

The solid fraction, coconut stearine, is a harder fat with a higher melting point. It is used as a valuable confectionery fat and as a substitute for cocoa butter because of its brittleness and "snap" fracture. The liquid fraction, coconut oleine, has a correspondingly

lower melting point and is used in margarine manufacture. Hydrogenation converts its unsaturated glycerides into stearic glycerides. The resulting product has a melting point higher than coconut stearine and is used as a brittle confectionery fat, which resembles cocoa butter. When refined and deodorized, coconut oil mixed with nonfat milk is often used as a replacement for whole milk. Other uses include the making of imitation dairy products, coffee whiteners, soft-serve desserts, frozen desserts, whipped toppings, milk shake mixes, and chocolate-filled milk.

Coconut oil is used because of its bland flavor, resistance to oxidation, and stability in storage, as well as a unique liquefying property that contributes to the "mouth-feel" of the food of which it is a component. The main nonedible uses are in soaps, detergent foam boosters, lubricating oil additives, mineral flotation agents, shampoo products, and corrosion inhibitors. Lauric oils enhance the lathering quality of soaps, and this quality makes coconut oil particularly useful for hard water or marine soaps. A feature of soapmaking with coconut oil is the higher yield of glycerol (14 percent compared with 10 percent for most oils). Other nonedible uses include illuminating or fuel oil in rural areas and in ceremonial lamps. Coconut stearine is also used to advantage in candle manufacture. Coconut oil will directly fuel unmodified diesel engines.

Copra

When the industrial demand for coconut oil developed in the nineteenth century, sailing schooners, and later tramp steamers, visited Pacific islands where the palm was plentiful. Fresh coconuts are a bulky and perishable cargo because of the husk and high water content. The fruit is made up of about 50 percent husk, 12 percent shell, 10 percent water, and only about 28 percent meat (kernel). The fresh coconut meat itself contains about 47 percent moisture. Thus, it was more convenient to ship copra, which the islanders could prepare in advance by drying the kernels, either in the sun or, if needed quickly, over a fire.

Commercial copra plantations today still use sun drying, direct firing over a barbecue, or indirect hot air in various sorts of kiln. The moisture content is reduced from between 45 and 50 percent to between 6 and 8 percent, and the oil content increases from 35 percent to between 60 and 65 percent. For safe storage, the moisture content of copra should be 6 percent, yet at the first point of sale it often has a much higher level. It dries further during storage, but molds may attack under such conditions. One of these is *Aspergillus flavus*, which produces aflatoxin. The possible presence of this carcinogen should serve as a stimulus for industry to improve copra quality or to bypass making it and process the fresh fruit instead (Harries 1993).

Edible (Ball) Copra

Copra may also form naturally inside the whole ripe nut. As early as the middle of the sixth century, Cosmas Indicopleutes said of the coconut: "If the fruit is gathered ripe and kept, then the water gradually turns solid on the shell, while the water left in the middle remains fluid, until of it also there is nothing left over" (Desmond 1992: 7). Copra formation occurs when nuts are kept in dry environments, and with those varieties that do not germinate quickly. The endosperm (kernel) eventually comes away from the shell and forms a ball of copra that rattles loosely inside the nut. The husk can remain intact so that the shell will not crack, and the whole copra-formation process requires some 8 to 12 months. Fires may be lit to help the drying. The heat and smoke do not contaminate the endosperm, which retains a very high quality (Harries 1993).

Copra Cake and Copra Meal

After the oil is extracted from copra, a good-quality residual cake will contain 6 to 8 percent oil, with a protein content of around 20 percent. Copra meal, the solvent-extracted residue, is 1 to 3 percent oil, depending on the efficiency of the plant. Cattle and poultry feeds incorporate both cake and meal; this combination results in firmer butter and harder body fat in cattle than that induced by other oil cakes. Cake with a high oil content is generally fed to pigs, but a deficiency in certain amino acids, notably tryptophan, lysine, methionine, and histidine, limits the amounts used in animal feeds. If aflatoxin is present in poorly prepared copra, it can pass into the cake or meal (Harries 1993).

Coconut Flour

Coconut flour suitable for human consumption is produced when oil is extracted from fresh coconut kernels rather than from copra. The flour is used in making bread and other foods. But, as just noted, it is not superior to other protein sources in the proportions of the various amino acids (Harries 1993).

Coconut-Confectionery

Sugar and Honey

The coconut palm is also a fine source of nectar. It begins to flower 3 to 5 years after planting, depending on growing conditions, and once started, it opens a new inflorescence regularly at 25- to 30-day intervals throughout the year. The palm goes on flowering for the remainder of its 80-year or longer life span. Every inflorescence includes hundreds of male flowers that open sequentially over a 3-week period, and each contains a drop of nectar when first open, which attracts honey bees and bumblebees in the early morning. Each inflorescence also carries female flowers, sometimes more than a hundred. For about a week each month, female flowers are receptive to pollination for a day or

two and produce an almost continuous flow of nectar droplets from three exposed and easily accessible nectaries. The flowers are also visited by birds (honeyeaters), and even by lizards.

The activity of insects draws attention to the nectar, whose sweetness is readily sampled by touching with a finger (easily reached in young palms). These may have been the clues that encouraged early domesticators and cultivators to find a way of increasing the flow of nectar. This method is known as tapping, and it produces toddy, as described in detail in the section "Water into Wine."

Sweet toddy, boiled in shallow pans to its crystallization point, gives a 12 to 15 percent yield of jaggery. This rough sugar is hard, semicrystalline, and golden brown in color. A lesser degree of concentration gives treacle (or syrup) (Harries 1993). Beehives are often kept in coconut groves to enhance fruit set. The year-round flowering in a coconut plantation assures a perpetual supply of nectar, and the hives can also serve as sources of pollen.

Desiccated Coconut

The characteristic taste of coconuts (when mature) and, to a certain extent, their texture (when grated and dried, or desiccated), are among their most important features for spicing and flavoring. In the United Kingdom, television advertisements for a chocolate-covered, coconut-filled confection give an entertaining, but faulty, impression that the coconut falls from the palm already peeled and neatly split in half. In reality, the manual labor involved in making desiccated coconut, including harvesting, peeling, cracking, deshelling, and shredding, is far from amusing and not necessarily rewarding. Australians like their favorite cake covered with desiccated coconut, which makes Australia a large importer of this product. Yet, farmers on neighboring Pacific islands, who cannot grow much else than coconuts, neglect the crop because of low world market prices for their product.

Desiccated coconut was first manufactured in the early 1880s. It is an important product, sensitive to changes in production costs, and easily susceptible to overproduction. Nuts are stored for three or four weeks before being dehusked in the field and carried to the factory. When the shell is chipped off, the kernel comes away easily. Damaged or germinated nuts are rejected to make low-grade copra. The brown testa is removed, usually by hand, though machines are available. The kernels are then washed and sterilized to avoid the risk of salmonella. After sterilization, disintegrators reduce them to a wet meal, or cutters produce fancy cuts, such as threads or chips. Drying is by indirect drier at 75–80° C, or by direct firing at 120° C. The dried product is cooled and graded before being packed. Parings, oil, and drain oil are byproducts.

Desiccated coconut should be pure white, crisp,

and have a fresh taste. It should have less than 2.5 percent moisture, 68 to 72 percent oil (on dry weight), less than 0.1 percent free fatty acid (as lauric), and about 6 percent protein. If there is more than 6 to 7 percent sucrose, then sugar has been added. Unavailable carbohydrate is about 18 percent and crude fiber about 4 percent, and there is some mineral content. Desiccated coconut is widely used in sweets, biscuits, cakes, and cake fillings (Harries 1993).

Coconut – Milk, Water, and Wine

The "Cocoa's Milky Bowl"

Dr. Samuel Johnson's *Dictionary* of 1755 contained run-together articles on coco (the nut) and cocoa (the source of chocolate). As a result, spelling became confused, and for some time the word coconut was misspelled as "cocoa-nut." Thus, the poetic allusion to the "cocoa's milky bowl" refers to the "coconut," *Cocos nucifera*, and not to cocoa, *Theobroma cacao* (Child 1974). Yet even explaining this commits a further solecism, because coconut "milk" is a manufactured product. Unfortunately the distinction between coconut milk and coconut water is not always kept clear, even in research publications by coconut scientists. Coconut milk is prepared by squeezing freshly grated endosperm, usually with a little added water, through cloth. On storing, coconut cream forms an upper layer, and when either emulsion is heated, a clear oil separates. This is the basis of the time-honored village method of oil extraction. But coconut cream is also produced industrially in both liquid and spray-dried forms, and the national cuisines of coconut-growing countries use it extensively (Harries 1993).

Water into Wine

Both alcoholic and nonalcoholic beverages that are products of the coconut palm depend on the technique known as toddy tapping. As with other fruit juices, the watery sap that is the toddy can be converted to wine and other products by fermentation and distillation (the sugar content of coconut water from the immature nut also allows it to be fermented, but this is not common). Many types of palms are tapped in Southeast Asia, and the practice dates from at least the seventh century (Burkill 1935).

Unlike maple or rubber trees, which are dicotyledons where the layer of cambium below the bark is tapped, palms are monocotyledons, and the vascular strands are scattered through the tissue. Casual observers sometimes think that it is the coconut leaf stalks that are tapped. In reality, tapping uses the unopened flowering inflorescence. This is a large structure which, when cut in the tapping process and seen from ground level, could, indeed, be mistaken for the cut leaf stalk. There are many flowering stalks within an inflorescence, each able to exude sap. They are packed tightly into an enveloping spathe that

would normally split to allow pollination. Binding the spathe tightly prevents it from splitting naturally. It may also be lightly beaten and flexed to stimulate sap flow. Once ready, the end is cut off to allow the sap to drip into a receptacle. The toddy tapper visits the palm, morning and evening, to decant the accumulated sap from the container before fermentation gets too active. Sap flow continues for many days, and each day a sliver is removed to reopen blocked vascular elements and increase flow. This continues until only a stump remains, and the next inflorescence in sequence is prepared.

Obviously, tapped bunches do not flower normally, and the palm ceases to set fruit. If the sap flow decreases, the palm is allowed to rest. The palm may respond with particularly high yields of fruit on the next normal bunches. Excessive tapping followed by high fruit set could shorten the life of the palm. However, the financial return to the farmer would more than compensate for this shorter life (Harries 1993).

Toddy is initially sweet and watery, and the containers used to collect it are rinsed but not sterilized between uses. Because the weather is warm where coconuts grow, and collection is slow because the palms have to be climbed, fermentation to alcohol is practically unavoidable.

Toddy produced overnight and collected first thing in the morning contains about 3 percent alcohol and 10 percent fermentable sugar. Certain additives may slow or stop fermentation. Otherwise, fermentation continuing for 33 hours produces palm wine with an 8 percent alcohol content. Sweet, unfermented toddy contains 16 to 30 milligrams of ascorbic acid per 100 grams, and the content changes little during fermentation. The yeast in fermented toddy adds vitamin B (Harries 1993).

Arrack is the product of the distilling of fermented toddy. Doubly distilled arrack is the basis of local gins, rums, and so forth, with the addition of appropriate flavors.

As with other wine-making substances, coconut toddy can also become vinegar. Fermenting toddy with free access to air produces 45 percent acetic acid in 10 to 14 weeks. This is matured in closed casks for up to 6 months and, perhaps, flavored with spices and colored with caramel.

Coconut Water

The entertainer Harry Belafonte may not have been completely accurate when he sang that coconut water was "good for your daughter" and "full of iron," or that it could "make you strong like a lion." But he was praising the one thing about coconut that makes it different from all other plants – the large amount of water in the immature fruit.

Modern texts on coconut underrate the value of coconut water or overlook the part it played in the domestication of the coconut. Earlier writers had no such reservations. In 1510, Ludovici de Varthema

wrote that "[w]hen the nut begins to grow, water begins to be produced within; and when the nut has arrived at perfection, it is full of water, so that there are some nuts which will contain four and five goblets of water, which water is a most excellent thing to drink . . ." (cited in Harries 1978). As mentioned, coconut water is often wrongly called milk. As early as 1583, by which time the coconut had become well known, Father Thomas Stevens praised the ubiquitous coconut and its refreshing milk [sic], saying "this is so abundant that after drinking the contents of one nut, you scarcely feel the need of another" (Desmond 1992: 29).

The immature fruit, used for drinking, will not fall naturally but must be cut from the palm. Bunches are selected just as they reach maximum size, when a jellylike endosperm begins to line the cavity of the still thin and soft shell. At this stage each nut is full size, full of water with no airspace (it does not splash when shaken), and very heavy. Usually, the harvester cuts one or two entire bunches of nuts and lowers them to the ground on a rope. If they fall, the weight of water cracks or even bursts the soft shell inside the soft husk, whereupon the water drains away and the fruit rots (Harries 1993).

The coconut that is freshly harvested from a bunch that has been in the sun has a natural effervescence and will hiss with released gas when opened. Nevertheless, nature's "packaging" of this "product" leaves it at a disadvantage against internationally trademarked colas and mineral waters, because young coconuts deteriorate over a few days unless kept cool. Cutting away some of the husk reduces their size so they can be more efficiently kept in refrigerated storage, which extends "shelf life" considerably. There are instances when coconuts meant for drinking are transported hundreds of kilometers in refrigerated trucks, but this occurs only when such a vehicle would otherwise travel empty, where the roads are good, and where an affluent urban market has no other access to coconut. Moreover, the use of the coconut as a drink is marginalized by most conventional agricultural treatments. It is seen as reducing the crop of copra from which oil is extracted.

At the proper stage, coconut water contains about 5 percent sugar, and a large nut may have as much as 25 grams of sugar. The water also contains minerals, amino acids, and vitamin C. In addition to fermenting easily, yielding alcohol and vinegar, coconut water has auxinic and growth-promoting properties when used in plant tissue culture. Historically, various medicinal values were attributed to it. There is no doubt that it is a fine oral rehydration fluid for the severe diarrhea of cholera and other diseases. Because coconut water is naturally sterile, it may be injected intravenously to substitute for blood plasma in emergency surgery, and in combination with egg yolk, it finds use as a diluent in artificial insemination.

Coconut – The Tree of Life

Depending on variety, coconut fruit takes from 11 to 15 months to reach maturity, and the palm produces a new inflorescence every 3 to 4 weeks. This means that all the stages of fruit development, from youngest to oldest, are present on any palm at any given time of the year. In the fourteenth century, Jordanus of Séveras, who thought that the coconut was a “marvel,” wrote that “both flowers and fruit are produced at the same time, beginning with the first month and going up gradually to the twelfth,” so that there are flowers and fruit in eleven stages of growth to be seen together (quoted in Desmond 1992: 9). In this respect it meets the specifications of the biblical Tree of Life, “which bare twelve manner of fruits, and yieldeth her fruit every month” (Revelations 22:2).

According to Peter Martyr (d’Anghiera), writing about 1552, “[s]ome people believe that the germs of these trees were brought by the waves from unknown regions” (Harries 1978: 270). Four hundred years or so later it is now speculated that coconut may have originated on the coasts and islands of Gondwanaland, after which a wild form floated into the Indian and Pacific Oceans, but not the Atlantic. Domestication subsequently occurred in the Malaysian region (Southeast Asia and the western Pacific). The wild and domestic forms were both taken into cultivation, and introgressive hybridization between them produced the wide range of varieties recognized today (Harries 1990).

The original importance of the coconut palm was to coastal communities. With fish and shellfish to eat, coconut provided refreshing, sweet, and uncontaminated drinking water in an otherwise saline environment. No tools were needed to get it, and daily consumption of the water contained in one or two coconuts was enough to ensure good kidney function. The wild type of coconut spread without human interference, but domestication enhanced its drinking qualities in particular. The domestic type depends on human activity for survival and dissemination (Harries 1979).

The coconut preceded the Polynesians in those parts of the Pacific region to which it could float, and the Polynesians took domesticated forms to the islands that they settled (Lepofsky, Harries, and Kellum 1992). Before the development of the copra industry, coconut was a multipurpose plant on small Pacific islands, and its food potential was neither more nor less important than any other use. But another use was in interisland transportation. The islanders in the Indian and Pacific Oceans discovered that coconut husk fibers could be important in building and rigging sailing ships. Moreover, they took young fruit on board, at the start of a voyage, as self-contained, individual servings of uncontaminated drinking water.

The coconut palm was first grown as a plantation crop in the 1840s, because the industrial process for making soap, patented in 1841, required a cheap source of oil, which coconut oil from copra (the dried endosperm of the nut) could provide. And then, between 1846 and 1867, the development of dynamite from nitroglycerine had the remarkable effect of turning glycerine, a once discarded by-product of soap manufacture, into a more profitable item (Harries 1978).

Thus, for industrial and political empire builders, the coconut was a cheap source of raw material and also of war material. The “coconut cult” and “coconut boom” were features of the stock market in the early years of the twentieth century, and coconut plantations were established throughout the tropics.

The strategic importance of the coconut following World War I was clearly demonstrated when the German territories in Africa and the Pacific, with their extensive plantations, were taken away as reparation. As a result, the Japanese administered the Caroline, Mariana, and Marshall islands and, in 1942, they added other important coconut-growing countries to their collection. At the time, Indonesia and the Philippines by themselves accounted for more than 50 percent of the world supply of copra; Indochina, Malaya, Borneo, New Guinea, the Solomons, and the Gilbert Islands provided a further 25 percent.

With the end of the World War II, as nuclear weapons displaced high explosives, the military importance of the coconut gave way to other oil crops. Thus, because of their high palmitic acid content, palm oil and cottonseed oil were preferred over coconut oil for making the new “conventional” weapon, napalm. Similarly, in industry, coconut oil soap, excellent for lathering in hard or saline water, and coconut fiber (coir), valued for resilient, water-resistant rope, were ousted by petroleum-based detergents and synthetic fibers (Harries 1978).

Other animal life forms besides humans have also been associated with the coconut. Two are of cultural interest in relation to coconuts, as well as being foods in their own right: the coconut crab and the palm weevil. The coconut crab, or “robber crab” (*Birgus latro*), is a massive land-living crab that can climb coconut palm stems and is reputed to cut off nuts before returning to the ground to eat them. Its association with the coconut is not purely fortuitous. The coconut travels long distances over the Indian and Pacific Oceans by interisland floating and can easily carry the small postlarval stages of the crab. This would account for the equally widespread distribution of an otherwise terrestrial crab, which only spends about 30 days of its larval life in coastal waters. Charles Darwin observed that the coconut crab “grows to a monstrous size,” and “is very good to eat” (Harries 1983). Unfortunately, on many islands

where it was once found, the crab has been eaten to extinction.

Palm weevils (*Rhynchophorus spp.*) are a serious pest in coconut groves, killing palms directly by burrowing in the stem, and indirectly as a vector of the red ring nematode. The palm weevil grubs grow as large as a man's thumb, and subsistence cultivators can collect hundreds of them from fallen or felled palm stems. When fried in their own fat and eaten, the larvae provide a protein- as well as an energy-rich diet.

Another insect activity related to the coconut is the gathering of pollen by honeybees, and today health-food shops sell coconut pollen. As with other pollens, it is collected by incorporating a trap in the hive entrance that removes the pollen pellets as the bees return from foraging. But it could also be collected directly from male flowers. Coconut breeders routinely harvest and process male flowers for kilogram quantities of pollen used in artificial pollination for F₁ hybrid seed production (Harries 1973). Here again, the year-round flowering of the coconut means that regular supplies of pollen are easy to maintain.

When coconut oil was first available in Europe, it was advertised as healthy, whereas animal fats or dairy products were associated with communicable diseases. Now noncommunicable diseases, such as heart diseases and cancer, are of more concern. But the routine use of coconut oil for frying fish or for making margarine had been discontinued in Western societies (mainly as a matter of economic supply and demand) long before the diet conscious became wary of coconut oil, and it continues to be used directly in tropical diets and for vegetable ghee in India.

Coconut oil is easily digested and is absorbed into the system almost as rapidly as butterfat. This is attributed to the low molecular weight of the fatty acids. In common with other vegetable oils, coconut oil contains virtually no cholesterol, but there are objections to its food use because of the high saturation of its fatty acids. In the United States, "tropical oils" have come under attack from pressure groups, whose criticisms overlook the fact that most coconut oil is used for nonedible purposes, and that many of its food uses are to improve the quality of factory-prepared products. Only in coconut-growing countries, where it makes lower quality protein and carbohydrates more acceptable and more digestible, is the coconut still used extensively for cooking. In fact, it may turn out that naturally saturated medium-chain coconut oil is healthier than artificially hydrogenated short-chain vegetable oils.

Finally, what more can be said about the coconut than was said in the "Account of Priest Joseph," circa 1505: "In conclusion, it is the most perfect tree that is found, to our knowledge" (cited in Harries 1978).

Hugh C. Harries

Bibliography

- Adair, D., and A. D. Marter. 1982. *The industrial production of coconut cream*. UNIDO Report 10528. Vienna.
- Adriano, F. T., and M. Manahan. 1931. The nutritive value of green, ripe and sport coconut (buko, niyog and macapuno). *Philippine Agriculturalist* 20: 3.
- Anon. 1912. *The cult of the coconut: A popular exposition of the coconut and oil palm industries*. London.
- Balakrishnamurthi, T. S. 1951. Food value of the coconut. *Ceylon Coconut Quarterly* 2: 113-14.
- Banzon, J., O. N. Gonzales, S. Y. de Leon, and P. C. Sanchez. 1990. *Coconuts as food*. Quezon City, Philippines.
- Behre, A. 1907. The composition of coconut milk. *Pharmazeutische Zeitung* 47: 1046.
- Blackburn, G. L., et al. 1989. A re-evaluation of coconut oil's effect on serum cholesterol and athero-genesis. *Asian and Pacific Coconut Community Quarterly Supplement* 18: 1-19.
- Blauvelt, K. 1939. The use of non-sterilized coconut milk as an additional nutrient substance in culture media. *Journal of Laboratory and Clinical Medicine* 24: 420-3.
- Bodmer, R. 1920. Desiccated coconut. *Analyst* 45: 18.
- Burkill, I. H. 1935. *A dictionary of the economic products of the Malay Peninsula*, Vol. 1. London.
- Child, R. 1936. Production of sugar from sweet coconut toddy. *Ceylon Trade Journal* 1: 410-15.
1937. Edible coconut oil. *Tropical Agriculturalist* 89: 270-80.
- 1940a. The food value of the coconut. *Journal of the Coconut Industry* (Ceylon) 3: 230-5.
- 1940b. A note on coconut flour. *Journal of the Coconut Industry* (Ceylon) 4: 117.
1941. Coconut toddy and products derived therefrom. *Young Ceylon* 10: 3.
1974. *Coconuts*. Second edition. London.
- Coconut Statistical Yearbook*. 1970. Asian and Pacific Coconut Community. Jakarta.
- Crawford, M. 1940. Coconut poonac as a food for livestock. *Tropical Agriculturalist* 94: 168-71.
- Cruz, A. C., and A. P. West. 1930. Water-white coconut oil and coconut flour. *Philippine Journal of Science* 41: 51-8.
- Damoderan, M. 1928. The fermentation of toddy and an account of the micro-organisms producing it. *Journal of the Indian Institute of Science* 11: 63-74.
- Date, A. 1965. *Desiccated coconut*. Tropical Products Institute Report No. G12. London.
- Dendy, D. A. V. 1975. Protein products from coconut. In *Food protein sources*, ed. N. W. Pirie, 43-6. Cambridge.
- Dendy, D. A. V., and B. E. Grimwood. 1973. Coconut processing for the production of coconut oil and coconut protein food and feed products. *Oleagineux* 23: 93-8.
- Desiccated coconut. 1886-7. *Tropical Agriculturalist* 6: 34, 771.
- Desmond, R. 1992. *The European discovery of the Indian flora*. New York.
- Fritz, R. 1983. *Die Gefässe aus Kokosnuss in Mitteleuropa 1250-1800*. Mainz am Rhein.
- Gibbs, H. D. 1911. The alcohol industry of the Philippine Islands. *Philippine Journal of Science* 6: 99-206.
- Gonzalez, O. N. 1986. State of the art: Coconut for food. *Coconuts Today* 4: 35-54.
- Grimwood, B. E., F. Ashman, D. A. V. Dendy, et al. 1975. *Coconut palm products: Their processing in developing countries*. FAO Agricultural Development Paper No. 99. Rome.

- Gwee Choon Nghee. 1988. *New technologies open the passage into new usage of coconut milk products*. Presentation at the Association of South East Asian Nations Food Conference '88. Kuala Lumpur.
- Hagenmaier, R. 1988. Fresh and preserved coconut milk. *Coconut Research and Development* 4: 40-7.
- Harries, H. C. 1973. Pollen collection from coconut flowers by means of a fluid bed dryer. *Euphytica* 22: 164-71.
1977. The Cape Verde region (1499 to 1549); The key to coconut culture in the Western Hemisphere? *Turrialba* 27: 227-31.
1978. The evolution, dissemination and classification of *Cocos nucifera*. *Botanical Review* 44: 265-320.
1979. Nuts to the Garden of Eden. *Principes* 23: 143-8.
1981. Milk bottle on the doorstep of mankind. Paper presented at the 13th International Botanical Congress, Sydney.
1983. The coconut palm, the robber crab and Charles Darwin: April Fool or a curious case of instinct? *Principes* 27: 131-7.
1988. Cod and chips - coconut and soap: Is there a connection? *Fish Friars Review* (October): 22.
1989. Coconut as a tropical fruit. In *Aspects of applied biology, Vol. 20: Tropical fruit - technical aspects of marketing*, ed. Association of Applied Biologists, 87-8. Wellesbourne, U.K.
1990. Malaysian origin for a domestic *Cocos nucifera*. In *The plant diversity of Malesia*, ed. P. Baas, K. Kalkman, and R. Geesink, 351-7. Leyden, Netherlands.
1991. Wild, domestic and cultivated coconuts. In *Coconut production: Present status and priorities for research*. World Bank Technical Paper No. 136, ed. A. H. Green, 137-46. Washington, D.C.
1993. Coconut palm. In *Encyclopaedia of food science, food technology & nutrition*, ed. R. Macrae, R. K. Robinson, and M. J. Sadler, Vol. 2: 1098-1104. London.
1994. Coconut. In *The evolution of crop plants*. Second edition, ed. N. W. Simmonds and J. Smartt, 389-94. London.
- Harries, H. C., and E. Almeida. 1993. Coconut as a fresh fruit. Poster Presentation at the November, 1993, International Symposium on Tropical Fruit, Vitória, Brazil.
- Hicking, A. 1949. Coconut milk: Substitute for dextrose in normal saline. *Hospital CPS Quarterly* (supplement to *Navy Medical Bulletin*) 22: 1-10.
- Jones, S. F. 1979. *The world market for desiccated coconut*. Tropical Products Institute Report No. G129. London.
- Leong, P. C. 1953. The nutritive value of coconut toddy. *British Journal of Nutrition* 7: 253-9.
- Lepofsky, D., H. C. Harries, and M. Kellum. 1992. Early coconuts in Mo'orea, French Polynesia. *Journal of the Polynesian Society* 101: 299-308.
- Levang, P. 1988. Coconut is also a sugar crop. *Oleagineux* 43: 159-64.
- Lin, F. M., and W. F. Wilkins. 1970. Volatile flavour components of coconut meat. *Journal of Food Science* 35: 538-9.
- Maciel, M. I., S. L. Olivera, and I. P. da Silva. 1992. Effects of different storage conditions on preservation of coconut (*Cocos nucifera*) water. *Journal of Food Processing & Preserving* 16: 13-22.
- Menezes, F. G. T., and B. N. Banerjee. 1945. Studies on the digestibility of edible oils and fats. 2. Effects of sterols, carotene and vitamins on pancreatic lipase. *Quarterly Journal of the Indian Institute of Science* 8: 7-30.
- Mohanadas, S. 1974. Preservation, bottling and keeping qualities of fresh coconut sap (sweet toddy). *Ceylon Coconut Journal* 25: 109-15.
- Monro, J. A., W. R. Harding, and C. E. Russell. 1985. Dietary fibre of coconut from a Pacific atoll: Soluble and insoluble components in relation to maturity. *Journal of the Science of Food & Agriculture* 36: 1013.
- Montenegro, H. M. 1985. Coconut oil and its by-products. *Journal of the American Oil Chemists Society* 62: 259.
- Moorjani, M. N. 1954. Milk substitute from coconut. *Bulletin of the Central Food Technology Research Institute (Mysore)* 4: 60-1.
- Naim, S. H., and A. Husin. 1986. Coconut palm sugar. In *Cocoa and coconut: Progress and outlook*, ed. E. Rajaratnam and Chew Poh Soon, 943-6. Kuala Lumpur.
- Nathaniel, W. R. N. 1952. The history of vinegar production and the use of coconut toddy as a raw material. *Ceylon Coconut Quarterly* 3: 83-7.
- Nicholls, L., and J. C. Drummond. 1945. Nutritive value of coconut. *Nature* 155: 392.
- Norris, K. V., B. Viswenath, and K. C. Nair. 1922. The improvement of the coconut jaggery industry of the West Coast. *Agricultural Journal of India* 17: 353-66.
- Northcutt, R. T. 1937. Light, fluffy, substantially non-rancidifiable coconut food product. *Chemistry Abstracts* 31: 4410.
- Ohler, J. 1984. *Coconut: Tree of life*. Rome.
- Pankajakshan, S. A. S. 1986. Report on the product testing of soft drink prepared from coconut water. *Indian Coconut Journal* 17: 3-10.
- Persley, G. J., M. A. Foale, and B. Wright. 1990. *Coconut cuisine*. Melbourne.
- Peters, F. E. 1952. The value of coconut meat as a human foodstuff. *Ceylon Coconut Quarterly* 3: 201-5.
1954. *Bibliography of the nutritional aspects of the coconut*. South Pacific Commission Technical Paper No 58. Noumea, New Caledonia.
- Phan, C. B., and R. R. del Rosario. 1983. The preparation of protein hydrolysate from defatted coconut and soybean meals. II. Quality and sensory evaluation of products. *Journal of Food Technology* 18: 163.
- Pinto, C. B. 1950. Study of coconut milk and its possible employment in therapeutics. *Revista de Societas Venezuelana Quimica* 4: 36-45.
- Ridley, H. N. 1912. *Spices*. London.
- Scheuring, J. J., and P. H. Tracy. 1942. Substitutes for cocoanut fat in dipping chocolate. *Ice Cream Review* 25: 12, 23, 38, 40.
- Som, M. N. M., et al. 1980. Processing of canned coconut milk and coconut butter. In *Cocoa & coconut: Progress and outlook*, ed. E. Pushparajah and Chew Poh Soon, 713-20. Kuala Lumpur.
- Subrahmanyam, V., and D. Swaminathan. 1959. Coconut as a food. *Coconut Bulletin* 13: 153-8.
- Swetman, A. A., and J. H. Broadbent. 1979. Sugar content variation of coconuts prior to the manufacture of desiccated coconut in Sri Lanka. *Tropical Science* 21: 33-8.
- Thampan, P. K. 1975. *The coconut palm and its products*. Cochin, India.
- The coconut as a food. 1883-4. *Tropical Agriculturalist* 3: 117.
- The milk in the coconut. 1883-4. *Tropical Agriculturalist* 3: 824.
- Thieme, J. G. 1968. *Coconut oil processing*. FAO Agricultural Development Paper No. 80. Rome.
- Thio, G. L. 1982. Small-scale and home processing of fresh coconut (oil manufacture) and utilization of by-products. *Bulletin of the Department of Agricultural Research* (Royal Tropical Institute, Amsterdam), 309-34.

- Timmins, W. H., and E. C. Kramer. 1977. The canning of coconut cream. *Philippine Journal of Coconut Studies* 2: 15-25.
- Unson, C. G. 1966. The effect of some food preservatives on coconut toddy. *Ceylon Coconut Planters' Review* 6: 22.
- Vanderbelt, J. M. 1945. Vitamin content of coconut milk. *Nature* 156: 174-5.
- Verghese, E. J. 1952. Food value of coconut products. *Indian Coconut Journal* 5: 119-29.
- Viswanath, B., and K. G. Mayer. 1924. Improvement of coconut jaggery industry. *Agricultural Journal of India* 19: 485-92.
- Woodroof, J. G. 1970. *Coconuts: Production, processing & products*. Westport, Conn.

II.E.3 🌴 Palm Oil

The oil palm (*Elaeis guineensis*) is a native of West Africa. It flourishes in the humid tropics in groves of varying density, mainly in the coastal belt between 10 degrees north latitude and 10 degrees south latitude. It is also found up to 20 degrees south latitude in Central and East Africa and Madagascar in isolated localities with a suitable rainfall. It grows on relatively open ground and, therefore, originally spread along the banks of rivers and later on land cleared by humans for long-fallow cultivation (Hartley 1988: 5-7).

The palm fruit develops in dense bunches weighing 10 kilograms (kg) or more and containing more than a thousand individual fruits similar in size to a small plum. Palm oil is obtained from the flesh of the fruit and probably formed part of the food supply of the indigenous populations long before recorded history. It may also have been traded overland, since archaeological evidence indicates that palm oil was most likely available in ancient Egypt. The excavation of an early tomb at Abydos, dated to 3000 B.C., yielded "a mass of several Kilograms still in the shape of the vessel which contained it" (Friedel 1897).

A sample of the tomb material was submitted to careful chemical analysis and found to consist mainly of palmitic acid, glycerol in the combined and free state, and a mixture of azelaic and pimelic acids. The latter compounds are normal oxidation products of fatty acids, and the analyst concluded that the original material was probably palm oil, partly hydrolyzed and oxidized during its long storage. In view of the rather large quantity found, the oil was probably intended for dietary purposes rather than as an unguent.

A few written records of the local food use of a palm oil (presumably from *Elaeis guineensis*) are available in accounts of European travelers to West Africa from the middle of the fifteenth century. Red palm oil later became an important item in the provisioning trade supplying the caravans and ships of the Atlantic slave trade, and it still remains a popular food-

stuff among people of African descent in the Bahia region of Brazil (Northrup 1978: 178-86; Hartley 1988: 1-3; R. Lago, personal communication, 1993).

The British Industrial Revolution created a demand for palm oil for candle making and as a lubricant for machinery. In the early nineteenth century, West African farmers began to supply a modest export trade, as well as producing palm oil for their own food needs. After 1900, European-run plantations were established in Central Africa and Southeast Asia, and the world trade in palm oil continued to grow slowly, reaching a level of 250,000 tonnes (metric tons) per annum by 1930 (Empire Marketing Board 1932: 117-23; Hartley 1988: 8-23; Lynn 1989: 227-31).

Meanwhile, the invention of the hydrogenation process for oils and fats in 1902 created the possibility of Western employment of palm products as, for example, in the making of margarine. Yet hydrogenation was more useful for liquid oils like groundnut, palm kernel, and coconut oils than for palm oil. After World War II, further improvements in palm oil refining technology and transport methods made it possible to use largely unhydrogenated palm oil in Western food products (Lim 1967: 130-2; Martin 1988: 45-8).

A rapid expansion of the palm oil export trade followed, accompanied by a marked growth in the plantation sector of production. Between 1962 and 1982, world exports of palm oil rose from about 500,000 to 2,400,000 million tonnes per annum, and Malaysia emerged as the world's largest producer, accounting for 56 percent of world production and 85 percent of world exports of palm oil in 1982. Expanded production in Malaysia was achieved mainly by the privately owned estate sector, which increased its oil palm holdings more than tenfold in the 1960s and 1970s; and by the Federal Land Development Authority (FELDA), whose large-scale schemes organized oil production along plantation lines, although ownership was vested in the workforce of "smallholders" (Khera 1976: 183-5; Moll 1987: 140-62).

By 1990, world production had reached nearly 11,000,000 tonnes per annum, with a worldwide trade of 8,500,000 tonnes (Mielke 1991: 110). Although red palm oil is still used in soups and baked dishes in West Africa, elsewhere in the world, palm oil is eaten mainly in a highly refined form. Its food uses vary from the vanaspati and ghee of India to the margarine, cooking oils, and biscuits of Europe and the United States.

The Oil Palm: Wild and Planted

West Africa

West Africa is the classic region of smallholder production, both of food and export crops. The oil palm, which has been both, flourishes in natural association with yam and cassava cultivation throughout the wetter parts of the region. In eastern Nigeria, which

C.W.S. Hartley (1988: 7, 16) called "the greatest grove area of Africa," densities of 200 palms per hectare (ha) were common in the late 1940s, and densities of more than 300 palms per hectare were not unknown.

These palms were typically self-seeded and tended (to varying degrees) by local farmers. Farther west, in the kingdom of Dahomey and in settlements established by the Krobo people near Accra, some deliberate plantings may have been made as the palm oil export trade developed from the 1830s (Manning 1982; Wilson 1991). However, as J. Reid (1986: 211) has noted, the word *plantation* was often used by contemporary European observers to mean a food farm on which oil palms happened to be growing. Moreover, although in Dahomey descriptions exist of seedlings being transplanted from the bush onto areas cleared for farming by slaves, this does not mean that the practice was universal. In any event, palm oil exports from Dahomey were much smaller than from the Niger Delta, where oil palms were planted deliberately in swampy regions outside their natural habitat, but where the bulk of production was carried out using natural groves. In the 1840s, Dahomey and the Niger Delta exported approximately 1,000 and 13,000 tonnes per annum respectively; by the 1880s these totals had risen to 5,000 and 20,000 (Manning 1982: 352; Reid 1986: 158; Martin 1988: 28-9; Lynn 1989: 241).

Beginning in the late nineteenth century, a number of experimental oil palm plantations were created by Europeans in West and west-central Africa. One of the earliest was founded in Gabon in 1870 by Roman Catholic missionaries (E. J. Mulder, personal communication, 1968). But like many of the other nineteenth-century plantations in West Africa, these ventures were unsuccessful. By comparison with African smallholders, European planters were highly specialized and vulnerable to the marked trading fluctuations of the late nineteenth century. Many also lacked capital but committed themselves to long-term investments of a type that African farmers sensibly rejected.

In the case of palm oil, money was spent paying laborers to produce, plant, and tend seedlings, often on marginal land, in regions where natural groves already contained more palms than local farmers could spare the time to harvest (Hopkins 1973: 212-14; Martin 1988: 46).

Thus, when in 1907 William Lever sought large-scale land concessions in the British West African colonies in order to produce palm oil for his Lancashire soap mills, the Colonial Office was reluctant to help him. In a region characterized by small, fragmented, and often communally owned farms, it was felt that Lever's scheme would be hard to administer, politically risky, and commercially unsound. Lever was left to pursue his dreams in the Belgian Congo, where the existing levels of both trade and population were far lower and where the colonial administration wel-

comed European enterprise (Hancock 1942: 188-200; Wilson 1954: 159-67).

Following the Lever debate, the West African palm oil industry remained in the smallholders' hands. Few other entrepreneurs came forward to press the case for plantations, although a number of state-run estates were established under French influence in the Ivory Coast after 1960. By 1981 these estates covered a total of 52,000 ha, with a further 33,000 ha planted with oil palms in the surrounding villages (Hartley 1988: 31).

Yet even this development was relatively modest in scale, as shown in unpublished data from Nigeria, West Africa's largest producer of palm oil. The area of wild palm groves, only partly harvested, was estimated at 2,400,000 ha, whereas there were 72,000 ha of estate plantations and another 97,000 ha of smallholder plantations. Estate plantations, which require large consolidated areas, are still difficult to create in Nigeria because the oil palm-growing regions are densely populated and the complex traditional land holding system has been carefully preserved. Elsewhere in West Africa, population densities are lower, but the problems of obtaining labor to sustain plantation developments are correspondingly greater.

Central Africa

In the late nineteenth century, both the German colonizers of Kamerun and the Belgian rulers of the Congo were keenly interested in applying European farming and processing techniques to the palm oil industry (Laurent 1912-13; Rudin 1938; Jewsiewicki 1983). But German botanical and mechanical trials were cut short by World War I, following which the German territories in Africa were divided between the French and the British.

In the Congo, however, Lever's initial land- and produce-buying concessions (granted in 1911) proved to be the foundation for a long process of experimentation, which eventually revolutionized the palm oil industry worldwide. New planting materials led to dramatic increases in yields, thus cutting the cost of production; and improved machinery led to high oil quality at a competitive price. Alongside developments in European and American food-processing techniques, the Congo innovations paved the way for the entry of palm oil into Western diets.

Lever was originally more interested in setting up mills than plantations in the Congo, but his initial investments brought heavy losses (Fieldhouse 1978: 507-9). The fruit supply from wild trees proved hard to control, both in the amount brought to the mill and in its quality upon arrival. Overripe or bruised palm fruit made for highly acidic, low-quality oil, whereas unripe bunches gave low yields. Yet Lever Brothers (and its successor Unilever after 1929) was unwilling to incur the heavy initial costs of planting trees unless planting materials were improved to reduce the running costs. The Germans in Kamerun had identified an exceptionally thin-shelled palm fruit with a high

oil content as early as 1902 (Hartley 1988: 50). But their "lisombe" palm, later to become known as the Tenera type, was found only rarely in the wild and failed to breed true.

In a renewed drive to encourage European investment in their colony and, in particular, in oil palm plantations, the Belgians began in 1922 to investigate this German discovery. An experimental plantation of Tenera palms was created at the Yangambi research station in the Congo, and in the 1930s these palms were subjected to a three-year testing program by M. Beirnaert. Meanwhile, private Tenera plantings had been made by Unilever and its subsidiary, the United Africa Company, in the British Cameroons and in the Belgian Congo itself (Courade 1977; Fieldhouse 1978).

Tenera seed also found its way to Sumatra and Malaya in the 1920s, although there, as in Central Africa, it failed to breed true. Beirnaert's painstaking experiments finally showed why: The Tenera palm was actually a hybrid of two other types, the thick-shelled Dura and shell-less Pisifera, and when self-pollinated would produce 50 percent Tenera, 25 percent Dura and 25 percent Pisifera (Beirnaert and Vanderweyen 1941).

Beirnaert's discovery was published at the height of World War II, and it was not until after 1945 that it could be turned to practical use with the establishment of large-scale and long-term Tenera breeding programs. It is ironic that the Congo was not the state that gained the most. Its oil palm plantations did expand from 52,000 ha in 1938 to 93,000 in 1945 and 147,000 in 1958, with a further 98,000 ha under smallholder cultivation by the end of that period (Mulder, personal communication, 1968; Hartley 1988: 30). But political unrest following independence in 1960 led to stagnation and decline in the industry.

Unilever, however, the most important single investor in 1960 with 47,000 ha under oil palms, remained loyal to the newly independent country of Zaire through two decades of intermittent losses and political uncertainty. Thus, Unilever managers remained in place following nationalization in 1975, and the company was allowed to take back full control of the estates two years later (Fieldhouse 1978: 530–45). But at a national level, the research effort was decimated, and new planting was very limited after 1960, in marked contrast to developments at the same time in Southeast Asia.

Southeast Asia

The oil palm was first introduced to Southeast Asia in 1848, when four seedlings, originating from West Africa, were planted in the botanical gardens at Buitenzorg (now Bogor) in Java (Hartley 1988, 21). But this introduction did not lead to a plantation industry for some time, although offspring of the palms were used as ornamentals by tobacco planters.

In 1905 a Belgian agricultural engineer, Adrien Hallet, arrived in Sumatra and noticed that its palms grew

more quickly and bore a richer fruit than counterparts in the Congo, where he had previously worked. Just as the oil palms in southeastern Nigeria bore a fruit with more oily pulp and a smaller kernel than their counterparts farther west, so did the Deli Dura palms, descended from the four Buitenzorg seedlings, hold a distinct advantage over the ordinary Duras of West and Central Africa (Leplae 1939: 25–7; Martin 1988: 47).

This superiority probably reflected the optimal soils, rainfall, and sunshine conditions of Southeast Asia, rather than any special genetic quirks of the Buitenzorg palms. However, the fact that all the Deli Duras were descended from so few parents meant that the early planters could expect fairly uniform results (Rosenquist 1986). This lowered the risks associated with plantation cultivation, an effect reinforced by the absence of the palm's usual pests and diseases in its new geographic setting.

The relatively high yields and low risks from planting oil palms in Southeast Asia helped the industry to grow quickly, following the pioneering plantings of Hallet in Sumatra and of his friend Henri Fauconnier in Malaya in the 1910s. By 1919 more than 6,000 ha had been planted in Sumatra, rising to 32,000 in 1925, by which time 3,400 ha had come under cultivation in Malaya. Over the next five years, a further 17,000 ha were planted in Malaya, while the Sumatran area doubled.

This rapid expansion came not only because of growing confidence in the oil palm but also because of the grave postwar problems of the rubber industry. The oil palm was seen as a useful means of diversification to avoid a dangerous dependence on rubber. The pace of new planting slowed during the worldwide slump of the 1930s, but by 1938 Malaya had nearly 30,000 and Sumatra more than 90,000 ha under cultivation (Lim 1967: Appendix 5.2; Creutzberg 1975: Table 11; Hartley 1988: 21).

Developments in Sumatra hung fire for some time after 1945, as shown in Table II.E.3.1. Meanwhile, developments in Malaysia were more rapid, especially after 1960, when the replanting of old rubber estates with oil palms was stimulated by FELDA's smallholder schemes.

Table II.E.3.1. *Indonesia: Oil palm area (thousand hectares)*

Date	Government schemes	Private estates	Small holders	Total
1940	–	–	–	110 ^a
1962	–	–	–	under 90 ^a
1971	90	50	–	140 ^b
1975	120	70	–	190 ^b
1980	200	90	10	300 ^b
1985	310	150	100	560 ^b
1990	–	–	–	1,100 ^c

Sources: ^aHartley (1988), pp. 21–7; ^bTaniputra et al. (1988); ^cBulletin Perkebunan (1991).

Table II.E.3.2. *Malaysia: Oil palm area (thousand hectares)*

Date	Peninsular Malaysia (Malaya)	Sabah	Sarawak	Total
1938	30	-	-	30 ^a
1960	55	-	-	55 ^a
1965	80	-	-	80 ^b
1970	261	38	1	300 ^c
1975	570	60	10	640 ^c
1980	910	90	20	1,020 ^c
1985	1,290	160	30	1,480 ^c
1990	1,660	270	60	1,990 ^c
1992	1,760	310	70	2,140 ^c

Sources: ^aHartley (1988) pp. 21-8; ^bKhera (1976) p. 28; ^cPalm Oil Registration and Licensing Authority.

At the same time, the Malaysian government and the estate sector launched several systematic Tenera-breeding efforts, in which high-yielding parents were selected and through which increasingly productive planting materials were generated. The new trees not only yielded more fruit but also produced a type of fruit that was ideally suited to the new screw presses which, having been tried out in the 1950s in the Belgian Congo, became widely used in Malaysia from the mid-1960s. These innovative developments have been described as "one of the world's outstanding agricultural achievements" (Anwar 1981). The land area involved is shown in Table II.E.3.2.

Latin America

A distinct species of the oil palm, *Elaeis oleifera* (also known as *Elaeis melanococca*), is indigenous to Latin America. Since the late 1960s, plant breeders have begun to take an interest in this variety because its oil has a high iodine value and unsaturated fatty acid content, making it especially suitable for food use. However, the fruit is often small, with a thin oil-yielding mesocarp surrounding a large, thick-shelled kernel. Harvested bunches often contain a low proportion of fruit of quite variable quality. Hence, the plant has not been cultivated commercially, although it is frequently found in riverside or swampy areas, and the oil is used locally for cooking, soap boiling, and lamp fuel (Hartley 1988: 85-9, 681-3).

Elaeis guineensis seeds were introduced to Central America by the United Fruit Company, which brought seeds from Sierra Leone to Guatemala in 1920, and from Malaysia to Panama in 1926 and Honduras in 1927. Other introductions from Java and the Belgian Congo followed, but the first commercial planting of 250 ha took place only in Guatemala in 1940. The United Fruit Company's main interest was, traditionally, the production of bananas for export, but large banana-producing areas were destroyed by Fusarium wilt, and in consequence, oil palm and other crops were being tested as replacements.

The oil palm, however, proved vulnerable to disease in its new setting, and difficulties were encountered in identifying suitable growing conditions. Nonetheless, a successful development was founded on the northern coastal plain of Honduras, and in addition to developing plantations on its own land, the United Fruit Company stimulated oil palm cultivation by neighboring smallholders, whose fruit could then be processed in the company mills. Seed was also supplied to other Latin American countries (Hartley 1988: 33-6; D. L. Richardson, personal communication, 1993).

The beginnings of commercial planting in Latin America are summarized in Table II.E.3.3. By 1992 the total area planted to *Elaeis guineensis* in Latin America had grown to 390,000 ha - still a small fraction of the area in Africa or Southeast Asia. The distribution of plantings by country and sector is shown in Table II.E.3.4.

Table II.E.3.3. *Latin America: Earliest oil palm plantings*

Country	Year	Area planted (hectares)
Guatemala	1940	250
Costa Rica	1963	638
Honduras	1964	447
Ecuador	1964	13
Brazil	1967	200
Colombia	1970	4,500
Mexico	1975	200
Venezuela	1976	560
Panama	1976	1,005
Peru	1978	5,500
Dominican Rep.	1981	160
Nicaragua	1984	600

Source: Richardson, personal communication (1993).

Table II.E.3.4. *Latin America: Oil palm area, 1992 (thousand hectares)*

Country	Estates	Smallholders (<100 ha each)	Total
Mexico	1	2	3
Nicaragua	3	-	3
Panama	4	-	4
Dominican Rep.	7	-	7
Guatemala	6	4	10
Peru	6	7	13
Venezuela	21	3	24
Costa Rica	16	9	25
Honduras	7	27	34
Brazil	31	28	59
Ecuador	59	29	88
Colombia	100	20	120
Total	261 (67%)	129 (33%)	390

Source: Richardson, personal communication (1993).

Processing Technology

Until the early years of the twentieth century, palm oil was processed only by traditional village methods, by which loose fruits were collected from the ground or a few bunches were cut from the tree. Beginning in the 1920s, however, the United Africa Company and British colonial officials in Nigeria started experimenting with steam cookers and hand presses designed to make production at the village level more efficient in terms of labor use and oil yield. Yet a lack of cash prevented most farmers from trying the new machinery, with the exception of a few lucky recipients of free samples or government subsidies in the 1940s (Martin 1988: 64–6, 127–9).

A separate process of trial and error led to the development of the sophisticated factories required to deal with the volume of fruit produced on modern plantations and to produce oil of the high and standardized quality that would appeal to Western food processors. Such factories handle almost all the palm fruit of Southeast Asia, whereas in West Africa and Latin America, processing is carried out by a wide variety of methods, yielding oil for local consumption and for industrial as well as edible uses in the West.

Whatever the scale and sophistication of the process, the following main steps are required:

1. Separation of individual fruits from the bunch.
2. Softening of the fruit flesh.
3. Pressing out of the oily liquid.
4. Purification of the oil.

Traditional Village Process

Whole, ripe, fresh fruit bunches (FFB) are cut from the palm. With young trees this can be done from ground level. With older trees in West Africa, harvesting is still often accomplished by a man climbing the tree, secured to it by a loop of rope or other locally available materials, such as rattan and raffia fiber (Vanderyst 1920). But on plantations, a curved knife attached to a bamboo is used. After cutting, most of the fruits are still firmly attached to the bunches, which are divided into a few sections, heaped together, moistened, and covered with leaves. Natural fermentation during two to three days loosens the fruits so that they can be picked off the bunch sections by hand.

Following this step, two major variants in the process are used to produce two oils with different characteristics – those of soft oil and those of hard oil. The regions producing each type have changed since Julius Lewkowitsch (1895: 429) identified Saltpond in present-day Ghana as the cheapest source of hard oil, and Drewin on the Ivory Coast as the best place to buy soft oil. But the basic methods have changed little since they were first described by

colonial officials in the 1910s and 1920s (Laurent 1912–13; Gray 1922; Faulkner and Lewis 1923; Martin 1988: 32–3).

For soft oil production, the fruits are separated as soon as they are loose enough and boiled with water for 4 hours to soften the flesh, which is very fibrous. The cooked fruit is emptied into a large container, which may be a pit lined with clay, an iron drum, or a large wooden mortar. It is then reduced to a pulp with pestles or by treading it under foot. The resulting mash may be diluted with water, and the oil is skimmed off or squeezed out of the fibrous mash by hand. In some instances, a sieve made of palm fronds is used to retain the fibers. At this stage the liquid product, which contains oil, water, and fruit fibers, is often boiled up with additional water and skimmed again, although this step is omitted in some cases. Finally, the oil is again heated to boil out the residual water.

Lewkowitsch (1922: 546) also reported on the preparation of small quantities of oil for kitchen use directly from freshly picked fruit, by boiling the fruit and skimming the oil. Such oil had good keeping properties and often a free fatty acid content below 2 percent; but yield was very low, and not available for export.

In the hard-oil process, the fruit is allowed to ferment for 3 or more days longer than in the soft-oil process, until the flesh is soft enough. It is then pulped by treading underfoot in an old canoe or pounding in a mortar. Oil is allowed to drain out for up to 3 days, then water is added, and the mix is trodden again. Further oil rises to the surface and is skimmed. The oil is boiled up with water in another container and finished as described for soft oil.

These two processes differ in some important respects. The prolonged fermentation in the hard-oil process results in a much greater enzymic breakdown of the neutral fat and, therefore, in a much higher free fatty acid content. The yield obtained by this process is also much lower. However, it has a substantial advantage in that the labor and firewood requirements are also much lower. Table II.E.3.5 summarizes these differences.

Table II.E.3.5. *Comparison of village palm oil processes*

	Soft oil	Hard oil
Proportion of oil recovered	40–50%	20–40%
Free fatty acid content	3–12%	20–40%
Elapsed time to produce clean oil (days)	5–7	10–14
Human days' work for one ton oil*	420	133

*Includes harvesting.

Sources: Nwanze (1965), and personal communication, J. H. Maycock of the Palm Oil Research Institute of Malaysia (1991).

The strong characteristic flavors developed during both of these processes, as well as the naturally strong red color of the oil, are appreciated by local cooks and visiting gourmets, but they present severe limitations in the export markets. The high free fatty acid content and solid consistency of hard oil limits its range of uses, making it well suited to soap boiling but not to food processing. The solid consistency of hard oil is not due directly to the free fatty acids formed during the fermentation step but rather to the diglycerides, the other fragments obtained when one fatty acid is split from the neutral triglyceride molecule. M. Loncin and B. Jacobsberg (1965) have demonstrated the formation of a eutectic between triglycerides and diglycerides, resulting in a minimum melting point and maximum softness at a diglyceride content of 7 percent.

Mechanization of the Small-Scale Process

With the rapid twentieth-century growth in West African exports came the introduction of simple machines to reduce labor requirements and increase oil yield from a given quantity of fruit. Early machines, before and after the 1914-18 war, as described by Hartley (1988: 694-703), included a cylinder fitted with manually operated beaters, which was fed with softened fruit and hot water. After "beating," an oil-water mixture was run off through a sieve. Another system used a special cooker and a press as adjuncts to the soft-oil process.

The first device to become widely adopted, however, was a modified wine and cider press: the Duchscher press. This consisted of a cylindrical cage of wooden slats, held in place by iron hoops, and a ram on a screw thread. The screw thread was turned manually by means of long bars (in the manner of a ship's capstan), forcing the ram onto the pulped fruit. The exuding liquid was collected in a trough surrounding the cage.

Similar presses, but using a perforated cylindrical metal cage, are still in use today, giving yields of 55 to 65 percent of the oil present. A recent analysis of the needs for mechanization in the village has concluded that this is still the most practical implement, because it can be made and maintained locally and is inexpensive by comparison with other presses (C. E. Williams, personal communication, 1981). However, farmers in Nigeria (which was once the world's largest exporter of palm oil) have, since the 1950s, been reluctant to invest in this or other improvements because of the low producer prices offered by the state-controlled marketing boards. It is to be hoped that recent reforms of marketing structures in Nigeria and elsewhere in Africa will encourage renewed innovation at the village level (Martin 1988: 126-36).

The next development in pressing was the introduction in 1959 of the hand-operated hydraulic press by Stork of Amsterdam. This was capable of process-

ing 600 to 1,000 pounds of fruit per hour and could recover 80 percent of the oil present. The hydraulic mechanism was later motorized.

A different approach to mechanization brought forth the Colin expeller (first patented in 1904), which in essence is similar to a domestic mincer. It consists of a perforated cylindrical cage, fitted with a spiral screw or "worm," which is turned manually through a gear. Cooked fruit is fed to the worm through a hopper, and the pressure developed as the worm pushes the fruit forward forces oil out through the perforations. Spent fiber and kernels are discharged at the end of the cage. The machine has a capacity of 100 kg cooked fruit per hour, or 250 kg per hour if motorized. The Colin expeller became popular after 1930, mainly in Cameroon. Its limitations were a reduced efficiency with Dura fruit, which forms the bulk of the wild oil palm crop; rapid wear of the screw; and a relatively high cost. The principle of the expeller, however, has been further developed into the screw press found in all modern oil mills (Hartley 1988: 703-5).

The presses described here provided a relatively efficient process for the step of pressing out the oily liquid during oil production and led researchers to seek improvements in the other steps. Several innovations have resulted from a project begun by the Nigerian Institute for Oil Palm Research (NIFOR) during the 1950s in cooperation with the Food and Agriculture Organization of the United Nations (FAO) and the United Nations Development Program (UNDP).

The following unit operations and equipment are involved in palm oil production.

1. Fruit bunch cookers, which are wood-fired cylindrical tanks. They are loaded with cut-up fresh fruit bunches (FFB).
2. A bunch stripper operated by hand, which consists of a cylinder made up of slats and turns on a horizontal axis that tumbles the cooked bunch sections until the individual fruits separate from the bunch and fall between the slats.
3. A digester (to break up the softened fruit and release its oil from its cells), consisting of a horizontal cylinder in which beater arms rotate, driven by a small diesel motor.
4. A hydraulic press, which was introduced in 1959.
5. A clarification unit consisting of two linked tanks, whereby heating with water causes the oil layer in the first tank to overflow at the top into the second tank. There it is dried by the waste heat from the fire under the first tank.

Extraction efficiencies of 87 percent at a free fatty acid (FFA) level of below 4 percent are routinely attainable by this process. Between a quarter and a half ton of fresh fruit bunches per hour can be processed, depending on cooker capacity.

A number of variants of this process are in use:

1. Bunches are allowed to ferment so that only loose fruit is loaded into the cooker. This variant yields oil of higher FFA.
2. The hydraulic presses may be driven by a small diesel engine.
3. Clarification can take place in simple tanks with direct heating.
4. Cooking of bunches may be by steam, whereby whole bunches are loaded into a tank fitted with a perforated plate about 15 centimeters (cm) from the base. Water is boiled under the plate, and the steam penetrates through the bunches.
5. In Ghana, an interesting operating procedure has been developed, in which the mill owner provides mill facilities to the farmers, who are then responsible for the bunch stripping and cooking of the fruit. Mill operatives carry out digesting and pressing procedures, after which the farmers take away the oil from their own fruit for clarification (G. Blaak, personal communication, 1989).

The advantages for the farmers are numerous: They need no capital investment in mill equipment; there are no arguments regarding purchase price of FFB; if farmers produce high-quality Tenera fruit, they retain the full benefit; and farmers pay only a processing charge. Their net profits are higher than those obtained selling FFB, even if they employ labor to carry out their share of the processing.

Larger-Scale Processes

The small-scale processes just described are suitable for the processing of FFB from wild oil palm groves or from smallholdings. The main objective is to produce red palm oil for traditional food use.

The processing of the large quantities of fruit produced by plantations or by large smallholder cooperatives, however, requires a progressively greater degree of mechanization and mechanical handling as the quantity increases. Furthermore, since oil produced on a large scale is usually intended for export or for local refinery processes, the ultimate objective is a neutral oil of bland flavor and nearly white color. To attain this quality, the processes (including the handling of FFB) are designed to minimize the development of free fatty acids and oil oxidation.

A simple factory process of intermediate scale, in which the material is still handled manually between processing stages, is the Pioneer mill, which was developed by the United Africa Company around 1939. It is designed to process about three-quarters of a ton of fruit per hour, which is the equivalent of about 1 ton of fruit bunches, following the removal of the fruit from the bunch stalks. The process consists of the following steps:

1. Autoclaving – 200 kg of fruit is loaded into a vertical batch autoclave mounted on a gantry and

cooked under steam pressure of 20 pounds per square inch for 15 minutes.

2. After cooking, the fruit is discharged by gravity into a digester fitted with a stirrer, which breaks it up and releases the oil from the cells.
3. The resulting mash is treated in a basket centrifuge, operating at 1,200 revolutions per minute.
4. The oil flowing from the centrifuge passes through a screen to remove the fiber, and then to a settling tank.
5. The settling tank contains a layer of hot water, and the oil is pumped in below water level. The water is boiled for 15 minutes and then allowed to settle. The oil layer is decanted through a hot-water layer in a second settling tank.
6. The tank is heated to boiling point for 15 minutes and allowed to settle. The clean oil is put into drums.
7. The sludge from the two settling tanks is further treated by boiling and settling, and the residual oil is recovered.

An oil mill of essentially the same design, with a capacity of 2 to 3 tons of fruit per hour, was featured in the Wembley Exhibition of 1924 by Nigerian Products Ltd., Liverpool, and was apparently demonstrated in operation there (Elsdon 1926: 316–22). In 1950 there were 13 Pioneer mills in operation in Nigeria. The numbers increased to 65 in 1953 and more than 200 in 1962, producing about 25,000 tons. But, subsequently, their use has declined (Mulder, personal communication, 1968).

The Pioneer mill cannot meet the needs of well-established plantations generating large volumes of fruit. To keep costs down and output up, it is vital to have a fully mechanized power-operated mill. The development of such mills began in Kamerun and in the Congo before World War I. Mills using centrifuges for oil extraction were in operation in the Congo in 1916, in Sumatra in 1921, and in Malaya in 1925 (Hartley 1988: 703–5). Centrifuges were largely replaced by hydraulic presses in the 1930s, although they were still being operated at Batang Berjuntai, Malaysia, in 1982. Batch-fed hydraulic presses were, in turn, replaced by continuous screw presses, which saved labor and handled much larger volumes of fruit. At this final stage of innovation, the development of agricultural and processing technology went hand in hand. The screw press tended to mangle the fruit of the Dura palm, with its relatively thin layer of oil-bearing mesocarp, but proved ideally suited to the Tenera variety (Maycock, personal communication, 1991).

The principal steps involved in the production of palm oil today are the following:

1. Harvest at optimum ripeness.
2. Transport FFB to an oil mill with minimum bruising.
3. Transfer FFB to sterilizing cages.
4. Sterilize FFB by steam under pressure.
5. Transfer cooked FFB to a bunch stripper.

6. Transfer fruit to a digester.
7. Press in single-screw or twin-screw press.
8. The oily discharge from the press, containing water and fruit debris, is passed through screens and settling tanks.
9. The oil phase from the settling tanks is passed to a clarifying centrifuge. The sludge, or heavy phase, from the settling tanks is centrifuged and the recovered oil returned to the settling tanks.

African Food Uses

In West Africa, palm oil has a wide range of applications. It is employed in soups and sauces, for frying, and as an ingredient in doughs made from the various customary starch foods, such as cassava, rice, plantains, yams, or beans. It is also a condiment or flavoring for bland dishes such as *fufu* (cassava). A basic dish, "palm soup," employs the whole fruit. The following dishes from Ghana are illustrative of the wide range of palm oil use (Wonkyi Appiah, personal communication, 1993).

In the case of palm soup, first wash and boil palm fruits. Next, pound the fruit and mix with water to a paste. Add meat or fish, vegetables, onions, spices, and salt. Boil for 25 minutes and simmer for a further 15. Serve the soup with cooked rice, yam, plantain, *fufu*, or *kpokpoi*. (The latter is a corn dough, steamed and cooked, with okra and palm oil stirred in.)

Palm oil is also used in baked dishes, and one popular dish has different names according to the local language. When *Ofam* in Twi, or *Bodongi* in Fanti, is prepared, ripe plantains are pounded and mixed with spices, some wheat or corn flour, beans, and perhaps eggs. Palm oil is stirred into the mixture and the whole is then baked in the oven. The dish is served with ground nuts. *Apiti* is a similar dish, baked in leaves, without beans or eggs.

The characteristic flavor of palm oil prepared by village methods is an important feature of these dishes. Indeed, it is one of their most "traditional" features. Several of the other key ingredients, such as salt, wheat, or (in popular eastern Nigerian dishes) stockfish, became widely available only in the nineteenth and early twentieth centuries, when they were imported from Europe in exchange for palm oil itself (Martin 1988: 28-9, 50).

Early Western Food Uses

The fully flavored red palm oil produced by West African village methods has not proved suitable for food use in the importing countries of the West, where the consumer requires a bland cooking fat, near white in color, or a margarine, similar in appearance to butter. Today's plantation-produced palm oil can be bleached and neutralized to meet Western requirements, but in the nineteenth and early twentieth centuries, the high FFA content even of "soft"

West African palm oils made them too difficult and uneconomic to neutralize (Andersen 1954: 27). Even before loading aboard ship, they fell far short of the current quality standard of less than 5 percent FFA, 0.5 percent moisture, and 0.5 percent dirt; and a slow ocean voyage did little to improve matters, as the acidity tended to increase en route (Vanneck and Loncin 1951).

Throughout the nineteenth century, exported oil from West Africa was placed in wooden casks usually supplied from Europe in the "knocked down" state and put together before being filled. Sailing ships became much larger in size and were gradually displaced in the second half of the century by steam ships, which were able to call at a greater number of ports and make more regular voyages (Lynn 1989). This development probably improved the overall quality of oil arriving in Europe, but as the oil was still made on a small scale by different methods and carried in casks, there was plenty of variation.

This quality problem could have been resolved in the late 1920s, when bulk storage tanks were installed at some African loading ports, initially in Nigeria and the Belgian Congo. It was then possible for incoming oil to be washed and cleaned before bulking, with an improvement in quality (Iwuchukwu 1965; Mulder, personal communication, 1968). However, hardly any Nigerian palm oil was suitable for the European food industry until the 1940s.

When Sumatran and, later, Malayan plantations started to export oil in the 1920s, their fruit was harvested systematically from the beginning. It was transported with minimal bruising to the factories and processed in a standardized way. Bulk shipment was developed from the outset. The first shore tanks were installed at Belawan in North Sumatra in the 1920s, and oil from Malaya was taken there by steamer from 1931 onward. In 1932 the Malayan planters set up their own Palm Oil Bulking Company with an installation at Singapore (Shipment of palm oil in bulk 1931; United Plantations Berhad, unpublished documents; T. Fleming, personal communication, 1993).

It thus became possible to develop and maintain the quality standards that are now current worldwide. The planters aimed to produce oil of 3.5 percent FFA, which would then fall well within the limit of 5 percent FFA on arrival in Europe or America. Oil arriving at above 5 percent FFA was sold at a discount, depending on the excess acidity (Hartley 1988: 687).

European food manufacturers could now begin to introduce palm oil on a commercial basis, drawing on earlier experiments and fitting it into two long-standing patterns of fat use. In central and northern Europe, indeed in cool weather regions generally, the traditional fats are the products of the farm yard - butter, beef tallow, and lard. In southern Europe, with its dry hot climate, olive oil has been in general use for thousands of years. Thus, consumers have had available either a plastic product of solid appearance

or a clear liquid oil, and the cooking and eating habits developed accordingly.

Respect for these traditions led to the invention in 1869 of margarine and its development as a replacement for butter, when the latter was in short supply. Margarine was originally made from beef fat, and the plastic nature of butter was attained by blending in a liquid fraction separated from beef fat by crystallization. Margarine proved so popular that European supplies of beef fat did not suffice. Imports from the New World were important in the nineteenth century, but various imported vegetable oils gradually took the place of beef fat margarine blends as refining techniques developed. The fact that even "soft" palm oil is a solid fat in temperate climates, with a consistency quite similar to butter, made it an obvious candidate for such experiments, and the first recorded trial took place in 1907 (Feron 1969).

The refining and bleaching process required to render suitable palm oil involved a great deal of research and empirical know-how. Illustrative is some unpublished correspondence (copies held by K. G. Berger) between Dr. Julius Lewkowitsch, a consultant chemist in oils and fats, and a Liverpool trading house, the African Association Limited. Dr. Lewkowitsch had invented a process for rendering palm oil into an edible product and had entered into an agreement, dated January 24, 1905, to share the costs of development of the process with the African Association.

Evidently the work proceeded rather slowly, because in September 1907, Lewkowitsch received a letter from the Vice-Chairman of the African Association, saying: "I have sent you under separate cover a sample of refined beef suet. . . . Would it be possible to have the samples of the palm oil products made up to appear like this sample? I am afraid I shall never satisfy my Co-Directors until I can show them a palm oil product they can eat." A successful prototype was probably produced eventually, because in 1910 a small manufacturing concern, V. B. Company, was incorporated and the African Association was paying Lewkowitsch a regular salary as managing director from 1910 to 1912.

The first decade of the twentieth century also saw the introduction of hydrogenation of oils, a process by which liquid oils could be turned into plastic or hard fats to a controlled degree. As a result, vegetable oil-based "shortenings" were produced to replace lard and beef tallow as ingredients for cakes, pastries, and biscuits and as frying fats. Once adequately refined, palm oil was easily introduced in blend formula for these types of products and had the advantage of not requiring hydrogenation. By the mid-1930s, the relatively clean and less acidic plantation-produced palm oil of Malaya and Sumatra was finding a ready market in the United States, where it was used not only in fine toilet soaps but also in the making of compound lard. Over 50,000 tonnes per annum were used in the

American compound lard industry between 1935 and 1939 (Lim 1967: 130-2).

Wartime interruptions of supplies from Asia during the 1940s forced American manufacturers to find substitutes for palm oil, and the market was slow to revive afterwards. However, in Britain, wartime shortages of butter encouraged the use of palm oil in both margarine and compound lard, and this market continued to grow in the 1950s (Lim 1967: 131). British manufacturers, through the home Ministry of Food and the West African Produce Control Board, were able to corner the market in British West African palm oil (Meredith 1986). The Produce Control Board and, from 1949, its successor, the Nigerian Oil Palm Produce Marketing Board, played an important role in bringing the quality of this oil up to the standards set in Southeast Asia.

A grading system was set up as follows:

Grade I	under 9 percent FFA
Grade II	9 to 18 percent FFA
Grade III	18 to 27 percent FFA
Grade IV	27 to 36 percent FFA
Grade V	over 36 percent FFA

Higher prices were paid for the better grades, and there was an immediate response from the village producers, which enabled a Special Grade palm oil to be specified in 1950 with maximum 4.5 percent FFA at time of purchase. A significant premium was paid for this oil, with the result that Special Grade oil, which was only 0.2 percent of production in 1950, jumped to over 50 percent by 1954. In 1955, the specification was tightened to 3.5 percent FFA, and by 1965 Iwuchukwu (1965) reported that more than 80 percent of material for export had reached this quality.

Market Developments Since 1950

The development (mainly since the 1950s) of convenience foods and of snack food manufacture on an industrial scale opened up additional uses for palm oil, because of its good resistance to oxidative deterioration and its better ability to withstand the high temperatures used in frying than most alternative oils (Kheiri 1987; Berger 1992). The market developed especially rapidly after 1970, as the trees planted during the 1960s in Malaysia matured and as the Malaysian government and estate sector began to promote their product more actively in the West. Asian markets had also become important by 1980, as Western processing techniques were applied to meet local needs.

Figure II.E.3.1 shows that the world production of palm oil, together with its share in world supplies of oils and fats, increased dramatically from 1970 onward. Yet by this time, in many parts of Africa the industry had declined (Table II.E.3.6). Nigeria, for example, had no exportable surplus of oil after 1970 and, in fact, became a net importer of palm olein in

the 1980s. Exports from Zaire became very limited in the 1980s, and the Ivory Coast, with exports of 60,000 to 100,000 tonnes per annum, was left as the only significant African supplier. Meanwhile, as shown in Table II.E.3.6, new Asian producers were emerging, in particular Papua New Guinea and Thailand. By 1990, exportable surpluses of 10,000 to 30,000 tonnes per annum were also reaching the world market from Honduras and Costa Rica (Mielke 1991: 111).

As Malaysian production grew, both the planters and the government realized that it was vital to improve processing methods and to encourage the growth in demand for palm oil. The estate sector took the lead during the 1960s, developing higher grades of crude palm oil to suit European needs. Later, the government joined in the development of refineries and new products to suit Asian, as well as Western, tastes.

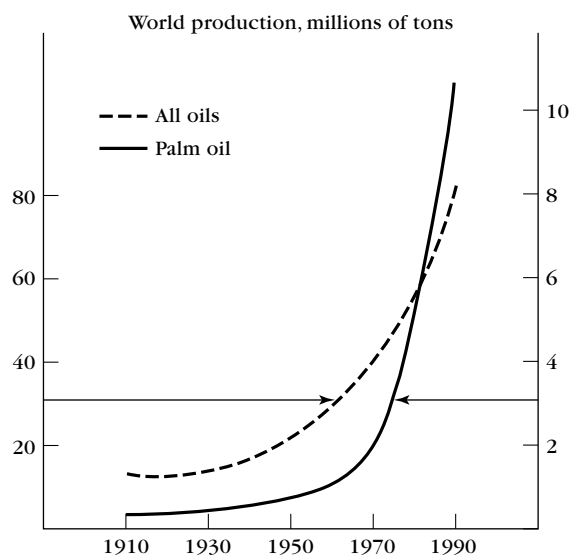


Figure II.E.3.1. World production of palm oil, 1910-90.

Table II.E.3.6. *Palm oil exports from selected countries (thousand tonnes per annum)*

Country	1950-3	1962-5	1970-3	1982-5	1990
Nigeria	180	130	8	-	-
Zaire	140	130	100	8	-
Indonesia	120	120	220	380	1,110
Malaysia	50	120	620	3,000	5,950
Ivory Coast	unknown	1	40	60	140
Papua New Guinea	-	-	-	100	130
Thailand	-	-	-	-	30
World total	520	540	1,000	3,700	8,440

Sources: 1950-85: Hartley (1988) p. 24, and 1990: Mielke (1991), p. 111.

Table II.E.3.7. *Specifications of special grades of crude palm oil*

	SPB	SQ	Lotox
FFA max %	2.0	2.0	2.5
Moisture max %	0.1	0.18	0.18
Impurities max %	0.002	0.02	0.02
Iron max ppm	10.0	4.0	3.0
Copper max ppm	0.5	0.2	0.2
Peroxide value mg/kg max	-	3.0	3.0
Anidine value max	-	-	4.0

Sources: Loncin and Jacobsberg (1965) and Johansson and Pehlgard (1977).

The old standard of 3.5 percent FFA on leaving the factory continued to apply to the bulk of crude Malaysian palm oil. However, in the last 30 years it has been recognized that the production of a stable refined palm oil of good color is also dependent on characteristics other than FFA content, as shown in Table II.E.3.7. In particular, the degrees of oxidation and contamination with catalytic traces of metals are important. But surveys indicate that the peroxide value (a measure of the state of oxidation) of standard palm oil arriving at Malaysian ports or refineries fell from 3.9 milligrams per kilogram to 2.3 milligrams per kilogram between 1974 and 1991 (Jacobsberg 1974; Wong 1981; V.K. Lal, personal communication, 1991).

Planters, both in the Belgian Congo (Zaire) and in Malaysia, also sought to develop a premium product with exceptionally low FFA, obtained through stricter harvesting routines and processing with minimum delay. "Special Prime Bleach" (SPB) grade was developed in the Belgian Congo (Loncin and Jacobsberg 1965), having a maximum of 2 percent FFA and reduced levels of iron and copper contamination, while in Malaysia two special grades became available, namely "SQ" and "Lotox." The SQ and Lotox specifications include limits for oxidation characteristics as well as trace metals and in practice satisfy the most stringent requirements of the European market (Johansson and Pehlgard 1977).

A separate development, which also improved the quality of palm oil arriving in Europe, was the introduction in the 1960s of "parcel tankers." These are specialized ships of up to 30,000 tons. The cargo space is subdivided into a number of separate tanks, generally with a capacity of between 100 and 1,000 tons. Tanks are fitted with separate pumps and pipelines so that different liquid cargoes can be carried without contamination. With the very large export trade from Southeast Asia, parcel tankers are capable of economically carrying palm oil and other oils of different grades to destinations all over the world. Appropriate shore installations have been developed since 1960 in the exporting ports of Southeast Asia and in most receiving ports in Europe, United States, and Japan, and are being developed in

countries that have only recently become large importers. Like the development of bulk shipment from Malaya in the 1930s, this innovation was fostered by cooperation among the planters, who marketed their oil through a common Malayan Palm Oil Pool from 1953 to 1974 (Allott and Wong 1977).

By 1974 the volume of Malaysian oil exports and the diversity of markets had grown to the point at which a free marketing system was more appropriate. The range of palm oil products exported was also growing, following the application of the fractionation process first developed for beef tallow in the 1870s. This technique separated crude palm oil into olein and stearin. The olein remains a liquid oil in hot climates and, therefore, readily fitted into the large Indian demand for cooking oil.

From 1970 the Malaysian government encouraged and licensed private enterprises to set up refineries that could both fractionate palm oil and use it in a more traditional manner to produce shortenings for Asian markets. In India, for example, there is a large consumer demand for vanaspati, a shortening developed as a replacement for butterfat ghee. Similar shortening products are traditional in Pakistan, Egypt, and other Middle East countries. Often lacking their own refining facilities, such countries have tended to import palm oil products from Malaysia in fully processed, ready-to-eat form.

The Malaysian government encouraged this trend by offering tax concessions for refineries in their early years and by progressive remission of the export tax on crude palm oil, graded according to the degree of processing of the end product. This development received a mixed reaction in Western Europe, which had ample processing capacity and extensive technical know-how (Berger, personal observation). However, it proved successful in stimulating the growth of new Asian markets, as shown in Table II.E.3.8.

Table II.E.3.8. *Palm oil imports to selected regions (thousand tonnes per annum)*

Region	1960	1970	1980	1990
W. Europe	470	520	830	1,600
United States	20	60	120	130
India	40	-	530	630
Japan	10	40	150	280
Former USSR	1	10	100	210
Pakistan	-	-	270	680
Egypt	-	-	-	330
Iraq	-	-	110	220
S. Korea	-	-	130	220
Turkey	-	-	-	200
Saudi Arabia	-	-	110	130
China	-	-	60	110
World total	620	930	3,682	8,440

Sources: Mielke (1988), pp. 76; (1991), p. 110; and Amiruddin and Ahmad (1984).

Although private enterprises had an excellent record in developing new processing techniques, they often felt hampered by their distance from major markets, which posed difficulties in designing and introducing new products. In 1979 the Malaysian government set up a specialized Palm Oil Research Institute (PORIM) as a statutory body, financed by a levy on palm oil production. A major part of its mission was the development of application technology for palm oil and the propagation of the information to end users anywhere in the world. By studying consumption patterns of oils and fats, the Institute's staff was able to identify potential new markets and provide the technical input needed for their development (PORIM 1981: 1-5). Their work proved useful to producers worldwide, especially from the late 1980s on when a debate arose over palm oil's nutritional value.

Nutritional Properties of Palm Oil

Briefly, the general nutritional functions of fat are to:

1. Provide energy efficiently.
2. Supply the essential linoleic and linolenic acids.
3. Carry the fat soluble vitamins A, D, and E.
4. Improve the palatability of foods.

The specific nutritional properties of palm oil may be considered in relation to its chemical composition. Typical values are given in Table II.E.3.9.

Table II.E.3.9. *Composition of pam oil*

1. The main fatty acids		Percentage	
C14:0		1.0	
C16:0		43.7	
C18:0		4.4	
C18:1		39.9	
C18:2		10.3	
C18:3		0.4	
Others		0.3	
2. The main glycerides		Percentage	Major components
Trisaturated		8.5	(PPP)
Mono unsaturated		37.7	(POP, PPO)
Di-unsaturated		11.7	(OOO, PLO)
More unsaturated		11.7	(OOO, PLO)
3. The nonglyceride components			
Crude palm oil	Ppm	Main component	
Carotenoids	640	beta-carotene, alpha-carotene	
Tocopherols	850	gamma-tocotrienol, alpha-tocotrienol, alpha-tocopherol	
Sterols	300	beta-sitosterol, campesterol	
Triterpene alcohols	800		

Source: Table based on Berger (1983a).

Table II.E.3.10. *Tocopherol content of typical refined palm oil*

alpha-tocopherol	180 ppm
alpha-tocotrienol	149 ppm
gamma-tocotrienol	239 ppm
delta-tocotrienol	62 ppm

Source: Gapor and Berger (1983).

The unsaturated acids present are mainly oleic, with a useful level of linoleic and a small amount of linolenic acid. In consequence, palm oil has a high stability to oxidation. Palm oil is readily absorbed and shows a digestibility of 97 percent or greater, similar to that of other common edible oils.

As in other vegetable oils, the middle 2-position is mainly occupied by unsaturated fatty acids. This is different from animal fats, where the 2-position is usually occupied by a saturated fatty acid.

Unrefined, or "virgin," palm oil is one of the richest natural sources of carotenoids. Regrettably, these are removed during the industrial refining process so that their nutritional benefits are lost, except to populations who traditionally use palm oil in the unrefined state.

The tocopherol content (see Table II.E.3.10) is one of the most interesting features in palm oil because it consists mainly of the tocotrienols, with an unsaturated side chain. These are not found in the other common vegetable oils.

Analytical work has shown that an average of 50 to 60 percent of the tocopherol content remains after refining, but the extent of removal depends on the refining conditions used. The tocopherols are important natural antioxidants, although their antioxidant activity is somewhat lower than the synthetic phenolic antioxidants permitted in foods. They are less volatile and, therefore, more persistent in high-temperature conditions, such as in deep-fat frying. The tocopherol content is a major factor in stabilizing palm oil against oxidation. The nutritional benefit of tocopherols in a number of disease conditions in which free radicals or oxidation are implicated has become a very active field of research, although to date little has been done on tocotrienols as such.

The three major component fatty acids of palm oil – palmitic, oleic, and linoleic acids – are also the most common fatty acids found in vegetable oils. Palm oil has been used as a traditional food in West Africa probably for thousands of years, which provides some evidence that it has good nutritional properties.

Research into coronary heart disease in relation to diet led to the general hypothesis of A. Keys, J. T. Anderson, and F. Grande (1957) that saturated fatty acids raised blood cholesterol levels, whereas linoleic acid reduced them. Subsequent refinements of the hypothesis (Hegstedt et al. 1965) indicated that saturated acids did not all have the same effect. In particular, D. M. Hegstedt and colleagues concluded that

myristic acid was 3 to 4 times more effective than palmitic acid in raising blood cholesterol levels.

The early work leading to these hypotheses did not use palm oil in the experimental diets. However, between 1985 and 1987, concern was expressed in the media, principally in the United States, that the saturated fatty acid content of palm oil meant that it would raise blood cholesterol levels and was, therefore, an undesirable food component. This assertion was not based on any direct experimental evidence. Instead, a review of the few dietary experiments in which palm oil had been used (as a control, not as the subject of investigation) showed that, in general, a small reduction in blood cholesterol level was experienced (New findings on palm oil 1987).

Subsequently, a study in Nigeria, a principal traditional consumer of palm oil, was published (Kesteloot et al. 1989). Serum lipid levels were measured in 307 men and 235 women, whose ages were 15 to 64 (mean 38.8) for men and 15 to 44 (mean 31.4) for women. Mean values for total serum cholesterol were 156.3 and 170.9, respectively, and for HDL cholesterol 46.0 and 49.0. Subjects consumed their normal diet, with 84 percent of the fat intake from palm oil. These serum lipid levels compared very favorably with black and white populations in the United States, where total fat intake is much higher, and where palm oil comprises only 1 to 2 percent of total fat intake (Park and Yetley 1990).

A number of new dietary studies have addressed the nutritional properties of palm oil, particularly in its effect on blood lipids. T. K. W. Ng and colleagues (1991), for example, found that when palm olein formed 75 percent of fat intake in a normal Malaysian diet, total serum cholesterol was significantly reduced, by 9 percent from the level at entry, and that the reduction was almost entirely in the undesirable LDL cholesterol. The study was carried out on 20 men and 7 women of average age 24.

A. Marzuki and colleagues (1991) used 110 residential high school students of both sexes as subjects (ages 16 to 17). Although the normal menu was provided, palm olein was the sole cooking oil for 5 weeks. This was followed by a "washout" period of 6 weeks on regular cooking oil and a second experimental 5 weeks in which only soya bean oil was used. There was no difference in plasma total LDL or HDL cholesterol between the two trial periods.

K. Sundram and colleagues (1992) carried out a double blind crossover study on 38 men, in which 70 percent of the fat in a normal Dutch diet was replaced by palm oil. There was no effect on total serum cholesterol, but a significant increase of 11 percent in HDL 2 cholesterol, resulting in a beneficial decrease in the LDL/HDL 2+3 ratio. G. Hornstra and co-workers (1991) also measured plasma lipoprotein (a), which is strongly associated with an increased risk of ischemic cardiovascular disease. They found a highly significant 10 percent decrease in this compo-

nent during the test diet period, and the decrease was greatest in subjects with an initial high level of lipoprotein (a), that is, those with an enhanced risk.

R. Wood and colleagues (1993) examined the effect of six different fats used as components of items of a normal American diet on 30 middle-aged men. When refined palm oil formed 60 percent of the dietary fat intake, there was no significant effect on total cholesterol, but HDL cholesterol was increased compared with the baseline diet.

Ng and colleagues (1992) studied 33 subjects consuming a Malaysian diet containing 34 percent of calories as fat. When palm olein was 23 percent of energy (that is, two-thirds of the fat intake), there was no significant difference in serum total - LDL or HDL cholesterol contents from the levels at entry. The use of olive oil in place of palm olein gave almost identical results, although the ratio of Thromboxane B₂ to Prostacyclin PGFI alpha was significantly lower during the palm olein dietary period.

D. Heber and colleagues (1992) found no increase in the plasma total cholesterol level of 9 subjects, but a small rise in plasma HDL cholesterol when one-half of the dietary fat was palm oil. The diet contained 35 percent energy as fat.

A. S. Truswell and co-workers (1992) conducted 2 trials (21 and 30 subjects, respectively) in which one-half of the dietary fat was palm olein. He found that a 10 percent increase in HDL cholesterol accounted for the 3 percent rise in total cholesterol observed.

The conclusion is that palm oil, used as a dietary fat at a high level - 10 to 20 times that usual in a Western diet - does not raise serum total cholesterol. However, the level of serum HDL cholesterol (popularly described as the "good" cholesterol, because in this form cholesterol is catabolized and removed) was significantly increased in several of the recent studies.

Mention might also be made of two additional studies, by R. C. Cottrell (1991) and C. E. Elson (1992). These authors reviewed 139 and 180 scientific publications, respectively, and both concluded that palm oil was a nutritionally satisfactory component of the diet. Cottrell wrote that "the decision to use palm oil in food products should be based on a rational appraisal of its technical performance value rather than on a misconceived view of the health implications of its use" (Cottrell 1991: 989S-1009S).

Nonetheless, the view still persists in some circles that palm oil is an unhealthy tropical grease, and it is difficult for palm oil producers to counter this perception because the product had little or no public image among Western and Asian consumers before the start of the recent media debate. Processed until it has become an anonymous ingredient, and used in a wide variety of compound fats and such other items as biscuits, its original flavor and feel have been lost to most consumers. But now that the wider nutritional benefits of palm oil's natural carotenoids are becoming more generally recognized, perhaps it is

time to rediscover the fully flavored red oil and promote its use, not only in Africa and Latin America but also in Asia and the West.

K. G. Berger
S. M. Martin

Bibliography

- Allott, D. H. N., and I. C. H. Wong. 1977. Evolution of palm oil marketing from Malaysia. In *International developments in palm oil*, ed. D. A. Earp and W. Newall, 469-78. Kuala Lumpur.
- Amiruddin, M. N., and I. Ahmad. 1984. Prospects for palm oil in the Middle East and selected African countries. In *Proceedings of the international seminar on market development for palm oil products*, ed. Y. Basiron and K. G. Berger, 131-9. Kuala Lumpur.
- Andersen, A. J. C. 1954. *Margarine*. London.
- Anwar, M. 1981. *Current and prospective situations of the oil palm/palm oil industry*. Palm Oil Research Institute of Malaysia, Occasional Paper No. 1. Kuala Lumpur.
- Beirnaert, A., and R. Vanderweyen. 1941. *Contribution à l'étude génétique et biométrique des variétés d'Elaeis guineensis jacquin*. Publications de l'Institut National pour l'Étude Agronomique, Série Scientifique 27. Belgian Congo.
- Berger, K. G. 1983a. Palm oil. In *Handbook of tropical foods*, ed. Harvey T. Chan and Marcel Dekker, 447-56. New York.
- 1983b. Past, present and future oils and fats supplies. *Proceedings of the Institute of Food Science and Technology* 16: 41-7.
1992. *Food uses of palm oil*. Kuala Lumpur.
- Bulletin Perkebunan. 1991. Vol. 22: 230-1.
- Cottrell, R. C. 1991. Introduction: Nutritional aspects of palm oil. *American Journal of Clinical Nutrition* 53: 989S-1009S.
- Courade, G. 1977. *Les plantations industrielles d'Unilever au Cameroun*. Yaounde, Cameroon.
- Creutzberg, P. 1975. *Changing economy in Indonesia*, Vol. 1. Amsterdam.
- Elsdon, G. D. 1926. *Edible oils and fats*. London.
- Elson, C. E. 1992. Tropical oils: Nutritional and scientific issues. *Critical Reviews in Food Science and Nutrition* 31: 79-102.
- Empire Marketing Board. 1932. No. 54: *Survey of vegetable oilseeds and oils*, Vol. 1: *Oil palm products*. London.
- Faulkner, O. T., and C. J. Lewis. 1923. Native methods of preparing palm oil. II. *Nigeria: Second Annual Bulletin of the Agricultural Department*, 3-10.
- Feron, R. 1969. Technology and production. In *Margarine: An economic, social and scientific history, 1869-1969*, ed. J. H. Van Stuyvenberg, 83-121. Liverpool.
- Fieldhouse, D. K. 1978. *Unilever overseas: The anatomy of a multinational, 1895-1965*. London and Stanford, Calif.
- Friedel, M. C. 1897. Sur des matières grasses trouvées dans des tombes égyptiennes d'Abydos. *Comptes Rendu* 24: 648-51.
- Gapor, A., and K. G. Berger. 1983. Tocopherols and tocotrienols in palm oil. In *Palm oil product technology in the eighties*, ed. E. Pushparajah and M. Rajadurai, 145-56. Kuala Lumpur.

- Gray, J. E. 1922. Native methods of preparing palm oil. *Nigeria: First Annual Bulletin of the Agricultural Department*, 29–50.
- Hancock, W. K. 1942. *A survey of British Commonwealth affairs*, Vol. 2, Part 2. London.
- Hartley, C. W. S. 1988. *The oil palm (Elaeis guineensis Jacq.)*. Third edition. London.
- Heber, D., J. M. Ashley, M. E. Solares, et al. 1992. The effects of a palm oil enriched diet on plasma lipids and lipoproteins in healthy young men. *Nutrition Research* 12: S53–S61.
- Hegsted, D. M., R. B. McGandy, M. L. Myers, and F. J. Stare. 1965. Quantitative effects of dietary fat on serum cholesterol in man. *American Journal of Clinical Nutrition* 17: 281–95.
- Hopkins, A. G. H. 1973. *An economic history of West Africa*. London.
- Hornstra, G., A. C. Van Houwelingen, A. D. M. Kester, and K. Sundram. 1991. A palm oil enriched diet lowers serum lipoprotein (a) in normocholesterolemic volunteers. *Atherosclerosis* 90: 91–3.
- Iwuchukwu, A. N. 1965. Influence of marketing boards system and preliminary purification on Nigerian palm oil exports. In *Proceedings of oil palm conference at Tropical Products Institute*, 73–7. London.
- Jacobsberg, B. 1974. Survey of the palm oil production in 1973. In *Proceedings of the first Malaysian Agricultural Research and Development Institute workshop on palm oil technology*, 72–119. Kuala Lumpur.
- Jewsiewicki, B. 1983. Rural society and the Belgian colonial economy. In *History of Central Africa*, Vol. 2, ed. D. Birmingham and P. M. Martin, 95–125. London.
- Johansson, G., and P. O. Pehlgard. 1977. Aspects on quality of palm oil. In *International developments in palm oil*, ed. D. A. Earp and W. Newall, 203–23. Kuala Lumpur.
- Kesteloot, H., V. O. Oviyasu, A. O. Obasohan, et al. 1989. Serum lipid and apolipoprotein levels in a Nigerian population sample. *Atherosclerosis* 78: 33–8.
- Keys, A., J. T. Anderson, and F. Grande. 1957. Prediction of serum cholesterol responses of man to changes in fats in the diet. *Lancet* 11: 959–66.
- Kheiri, S. A. 1987. End uses of palm oil: Human food. In *Critical reports on applied chemistry*, Vol. 15: *Palm oil*, ed. F. D. Gunstone, 71–83. London.
- Khera, H. S. 1976. *The oil palm industry of Malaysia: An economic study*. Kuala Lumpur.
- Laurent, M. 1912–13. Notes sur l'Elaeis au Congo Belge. *Bulletin Agricole du Congo Belge*, 3: 938–45; 4: 695–701.
- Leplae, E. 1939. *Le palmier à huile en Afrique: Son exploitation au Congo Belge et en Extrême-Orient*. Brussels.
- Lewkowitsch, J. 1895. *Chemical technology and analysis of oils, fats and waxes*. London.
1922. *Chemical technology and analysis of oils, fats and waxes*. Sixth edition. London.
- Lim, C.-Y. 1967. *Economic development of modern Malaya*. Kuala Lumpur.
- Loncin, M., and B. Jacobsberg. 1965. Research on palm oil in Belgium and the Congo. In *Proceedings of oil palm conference at Tropical Products Institute*, 85–95. London.
- Lynn, M. 1989. From sail to steam: The impact of the steamship services on the British palm oil trade with West Africa, 1850–1890. *Journal of African History* 30: 227–45.
- Manning, P. 1982. *Slavery, colonialism and economic growth in Dahomey, 1640–1960*. Cambridge.
- Martin, S. M. 1988. *Palm oil and protest: An economic history of the Ngwa region, south-eastern Nigeria, 1800–1980*. Cambridge.
- Marzuki, A., F. Arshad, T. A. Razak, and K. Jaarin. 1991. Influence of dietary fat on plasma lipid profiles of Malaysia adolescents. *American Journal of Clinical Nutrition* 53: 1010S–14S.
- Meredith, D. 1986. State controlled marketing and economic “development”: The case of West African produce during the Second World War. *Economic History Review*, 2nd series 39: 77–91.
- Mielke, S., ed. 1988. *Oil world*. Hamburg.
1991. Commodity section. *Oil world annual*. Hamburg.
- Moll, H. A. J. 1987. *The economics of the oil palm*. Wageningen, Netherlands.
- New findings on palm oil. 1987. *Nutritional Review* 45: 205–7.
- Ng, T. K. W., K. Hassan, J. B. Lim, et al. 1991. Non-hypercholesterolemic effects of a palm oil diet in Malaysian volunteers. *American Journal of Clinical Nutrition* 53: 1015S–20S.
- Ng, T. K. W., K. C. Hayes, G. DeWitt, et al. 1992. Dietary palmitic acid (16:0) and oleic acid (18:1) exert similar effects on serum cholesterol and lipoprotein profiles in normo cholesterolemic humans. *Journal of the American College of Nutrition* 11: 383–90.
- Northrup, D. 1978. *Trade without rulers: Pre-colonial economic development in south-eastern Nigeria*. Oxford.
- Nwanze, S. C. 1965. Semi-commercial scale palm oil processing. *Proceedings of oil palm conference at Tropical Products Institute*, 63–6. London.
- Palm Oil Registration and Licensing Authority. 1993. *Palm Oil Update: March*. Kuala Lumpur.
- Park, Y. K., and E. A. Yetley. 1990. Trend changes in use and current intakes of tropical oils in the United States. *American Journal of Clinical Nutrition* 51: 738–48.
- PORIM. 1981. *Palm Oil Research Institute of Malaysia: Annual research report*. Kuala Lumpur.
- Reid, J. 1986. Warrior aristocrats in crisis: The political effects of the transition from the slave trade to palm oil commerce in the nineteenth century kingdom of Dahomey. Ph.D. thesis, University of Stirling, Scotland.
- Rosenquist, E. 1986. The genetic base of oil palm breeding populations. *Proceedings of international workshop on oil palm germplasm and utilisation, March 1985*. Kuala Lumpur.
- Rudin, H. R. 1938. *Germans in the Cameroons, 1884–1914: A case study in modern imperialism*. New Haven, Conn.
- Shipment of palm oil in bulk. 1931. *The Planter* 11–12: 353–4.
- Sundram, K., G. Hornstra, A. C. Van Houwelingen, and A. D. M. Kester. 1992. Replacement of dietary fat with palm oil: Effect on human serum lipids, lipoproteins and apolipoproteins. *British Journal of Nutrition* 68: 677–92.
- Taniputra, B., A. W. Lubis, K. Pamin, and S. Syukur. 1988. Progress of oil palm industry in Indonesia in the last 15 years: 1971–1985. *Proceedings of international oil palm conference, Kuala Lumpur, September 1987*, 27–34. Kuala Lumpur.
- Truswell, A. S., N. Choudhury, and D. C. K. Roberts. 1992. Double blind comparison of plasma lipids in healthy subjects, eating potato crisps fried in palm olein or canola oil. *Nutrition Research* 12: S43–S52.
- Vanderyst, R. P. H. 1920. La technique indigène de récolte des fruits de l'Elaeis. *Bulletin agricole du Congo Belge* 11: 22–35.
- Vanneck, C., and M. Loncin. 1951. Considérations sur l'alter-

- ation de l'huile de palme. *Bulletin agricole du Congo Belge* 42: 57-64.
- Wilson, C. 1954. *The History of Unilever*, Vol 1. London.
- Wilson, L. E. 1991. *The Krobo people of Ghana to 1892: A political and social history*. Athens, Ohio.
- Wong, C. Y. 1981. Malaysian crude palm oil quality delivered to refiners. In *Palm oil product technology in the eighties*, ed. E. Pushparajah and M. Rajadurai, 81-7. Kuala Lumpur.
- Wood, R., K. Kubena, R. Crook, et al. 1993. Effects of palm oil, margarine, butter, and sunflower oil on the serum lipids and lipoproteins of hypercholesterolemic middle aged men. *Journal of Nutritional Biochemistry* 4: 286-97.

II.E.4 Sesame

Botanical Description

Sesame (*Sesamum indicum* L.) belongs to the Pedaliaceae, a small family of about 15 genera and 60 species of annual and perennial herbs. These occur mainly in the Old World tropics and subtropics, with the greatest number in Africa (Purseglove 1968). Sesame is a crop of hot, dry climates, grown for its oil- and protein-rich seeds. The oil is valued for its stability, color, nutty flavor, and resistance to rancidity.

A large number of cultivars are known (Bedigian, Smyth, and Harlan 1986). These differ in their maturation time, degree of branching, leaf shape and color, and number of flowers per leaf axil, which may be 1 or 3. The locules in the capsule usually number 4 or 8. The known cultivars also vary in length of capsule, in intensity of flower color, and especially in seed color, which ranges from pure white to black, with intervening shades of ivory, beige, tan, yellow, brown, red, and gray. The seeds are about 3 millimeters long and have a flattened pear shape. The capsules open automatically when dry, causing the seed to scatter.

Production

Sesame is usually grown as a rain-fed crop. It has many agricultural advantages: It sets seed and yields relatively well under high temperatures, it is tolerant of drought, and it does reasonably well on poor soils. It is very sensitive to day length and is intolerant of waterlogging. The major obstacle to the expansion of sesame is its habit of shattering: The absence of non-shattering cultivars, suitable for machine harvest, results in labor-intensive harvest seasons. Because of this obstacle, the crop is not suitable for large-scale commercial production (although breeding for non-shattering traits has been ongoing). Instead, sesame has typically been grown on a small scale for local consumption or in places where labor is cheap. The

major sesame-producing countries, listed in order of decreasing tonnage, are India, China, Myanmar (formerly Burma), Sudan, Nigeria, and Uganda (FAO 1992).

Chemical Composition and Nutritional Value

Oil

The chief constituent of sesame seed is oil, which usually constitutes 40 to 60 percent of seed weight. This is a greater oil content than most other oil crops (Hussein and Noaman 1976; Tinay, Khattab, and Khidir 1976; Salunkhe et al. 1991). Sesame oil content is related to seed color, although this is specific to geographic location. For example, the white-seeded varieties in the Sudan have a lower oil content than the red-seeded varieties (Bedigian and Harlan 1983), but in Ethiopia white or light-colored seeds usually have more oil than dark-colored seeds (Seegeler 1989).

Seed color differences are also economically significant. In West Bengal, white-seeded sesame is sold at a price at least 30 percent higher than that of brown-seeded or black-seeded varieties because of its higher oil content and greater culinary utility (Chakraborty, Maiti, and Chatterjee 1984). But appreciation of sesame by seed color can also be cultural: The Japanese, for example, prefer black-seeded varieties (Y. Kakehashi, personal communication, 1995), whereas the Sudanese prefer white-seeded ones.



Sesame

In a collection of sesame plants from around the world, tested by D. M. Yermanos and colleagues (1972), the mean oil content in 721 samples was 53.1 percent; the oil was clear and colorless in 47.4 percent of the samples and light green in 37.2 percent, with the remaining samples dark green or brown. Short plants tended to have colorless oil, whereas tall plants had light green oil, and early plants had higher seed oil content. Earliness, yellow seed color, and large seed size were associated with a lower iodination value. However, Y. K. Sharma, A. D. Deodhar, and S. K. Gangrade (1975) reported no relationship between oil content and iodination values, free fatty acids, and carbohydrate content.

Fatty acids. Sesame oil contains about 80 percent unsaturated fatty acids, with oleic acid and linoleic acid predominating and present in approximately equal amounts (Lyon 1972). Sesame has more unsaturated fatty acids than many other vegetable oils, and its higher proportion of unsaturated to saturated fatty acids makes it a potentially important dietary source of essential fatty acids. Linoleic acid is required for cell membranes, for transportation of cholesterol in the bloodstream, and for blood clotting (reviewed in Salunkhe et al. 1991).

Antioxidants. Sesame oil is remarkably stable because of its natural antioxidants, sesamin and sesamol, two compounds not found in any other oil (Bedigian, Seigler, and Harlan 1985). These are phenylpropanoid lignans and serve to protect against rancidity (Weiss 1971; Salih 1993), which may be one reason that sesame is nicknamed the "Queen of Oilseeds." Other unsaponifiable substances in sesame oil include sterols, triterpenes, pigments, and tocopherols.

Protein

The press cake remaining after the oil is expressed is approximately half the initial seed weight and high in protein. C. K. Lyon (1972) reported a protein content range of 17 to 32 percent. Although low in lysine, sesame proteins are especially rich in the essential amino acids methionine and tryptophan, and sesame meal flour is recommended as an excellent source of methionine-rich proteins (Villegas, González, and Calderón 1968). These can be a valuable supplement to pulse proteins, such as those in beans or chickpeas, which contain adequate amounts of lysine but are usually deficient in sulfur-containing amino acids (Fernández de Campoy 1981). S. A. Godfrey, B. J. Frances, and C. S. Kamara (1976) have suggested a combination of cowpea (deficient in methionine) and sesame to overcome the problem of limiting amino acids.

Carbohydrates

The carbohydrate content of sesame seeds (21 to 25 percent) is comparable to that of peanuts and higher than that of soybeans (Joshi 1961). D. J. Wankhede and

R. N. Tharanathan (1976) have reported that sesame contains 5 percent sugars – d-glucose, d-galactose, d-fructose, sucrose, raffinose, stachyose, planteose, and sesamose – as well as higher homologues of the planteose series.

Minerals and Vitamins

Sesame seeds are a good source of calcium, phosphorus, and iron. Nearly one-half to two-thirds of the calcium in the seed is present as oxalate, and a major part of it is located in the outer husk of the seed. The seeds are often dehulled for food use; thus, the oxalate does not interfere with the absorption of calcium. Sesame seeds contain high levels of thiamine, riboflavin, and nicotinic acid (Tabekhia and Mohammed 1971). When made into sesame butter, they lose 52.5 percent of their thiamine and 50.2 percent of their nicotinic acid. The oil is rich in vitamin E.

Minor Constituents

Sesame contains among the highest levels of phytate found in nature (Boland, Garner, and O'Dell 1975). In foods, phytate can be detrimental to the absorption of trace elements such as zinc, presumably through formation of a stable insoluble complex that does not release zinc to the absorption sites in the intestinal mucosa. It is conceivable that the indigestible phytate-protein complexes could make zinc and other nutrients even less biologically available. In attempting to isolate and characterize the phytate, B. L. O'Dell and A. de Boland (1976) found that it was not readily extracted by neutral aqueous solvents because it exists as an insoluble magnesium complex.

While studying the natural occurrence of *Fusarium* toxins in feedstuffs, C. J. Mirocha and colleagues (1976) reported, for the first time, the natural occurrence of zearalenone in sesame seed. The presence of aflatoxin B₁ was sought in 292 samples of seeds other than peanuts, and sesame was among the 8 found to contain it, at 0.06 mg per kg (Monacelli et al. 1976).

In addition, unhulled sesame contains 1 to 2 percent oxalic acid (Lyon 1972), which can be removed by decortication.

Changes Effected by Processing

Roasting

Sesame seed is used as a nourishing food as well as a flavoring agent, and its characteristic flavor is developed by dry-roasting the dehulled material. S. N. Chaudhry (1988) investigated the chemical changes in protein quality induced by roasting at various temperatures. He determined that optimal roasting of sesame (with respect to flavor and oil quality) required a temperature of 200° C for a duration of one hour. G.-C. Yen and S.-L. Shyu (1989), however, concluded that sesame oil prepared under optimum processing conditions meant roasting at 200° C for 30 minutes.

Extraction

Sesame oil can be extracted from the whole seeds by several different processes (Johnson and Raymond 1964). Chaudhry (1988), who compared press-extracted to solvent-extracted oil (from roasted sesame seeds), reported that press extraction can release impurities (such as pigments and metal ions) that could react with the oil, resulting in complex formation. Solvent extraction, on the other hand, can cause the loss of desirable flavoring compounds during the evaporation of the solvent. A method of crushing sesame seeds for oil extraction is described in *Law's Grocer's Manual*, published about 1892:

Sesame is also widely cultivated in Syria, where, in preparing the oil, the grain is soaked in water for 24 hrs, and then placed in an oblong pot, coated with cement, on which two men work a wooden hammer of 20 lb. weight. . . . Efforts are not made to mash the kernels. The skins are separated in a tub of water, salted to a degree sufficient to float an egg. The bran sinks, while the kernels remain on the surface. The sesame seeds are now broiled in an oven, and sent to the mill to be ground. From the millstone the oil drops into a jar, and is thick, of a dark yellow colour, and sweet (quoted in David 1970: 54).

Toxicity

Sesame seeds are extremely powerful allergens, and sensitized persons may suffer urticaria, Quinke's edema, asthma, and even anaphylaxis (Kägi and Wüthrich 1993). Although such reactions to sesame seeds are rare, the potential danger should be recognized, especially in view of an increasing demand for vegetarian foods.

Origin

Sesame was cultivated in ancient India, Sumer, Egypt, and Anatolia and throughout the Greco-Roman world for its edible seed and for its oil (Bedigian 1985; Bedigian and Harlan 1986). It is often described as the oldest oilseed plant used by humans (Joshi 1961; Weiss 1971). There has been some confusion about its origin, however, that has been discussed by N. M. Nayar and K. L. Mehra (1970), Nayar (1976), and Dorothea Bedigian (1984).

A group of wild and weedy forms native to India and described as *Sesamum orientale* L. var. *malabaricum* Nar. shows close morphological, genetic, and phytochemical affinities to the cultivar. Bedigian (1988) and colleagues (Bedigian, Seigler, and Harlan 1985; Bedigian, Smyth, and Harlan 1986) have provided evidence to support the theory that domesticated sesame arose from this progenitor on the Indian subcontinent. D. Zohary and M. Hopf (1994) concurred with Bedigian and J. R. Harlan (1986) that the botanical evidence supports a relatively late introduc-

tion of sesame into the Near East and the Mediterranean region. The genus is restricted to the Indian subcontinent, where there are only a few wild species, and to Africa south of the Sahara, where there are numerous wild species.

The Archaeological Record

The Indian Subcontinent: Harappa

The oldest known remains of sesame seeds were found at the Indus Valley civilization site of Harappa (Vats 1940), in Pakistan, where excavators uncovered "a quantity of lumped and burnt *Sesamum*" specimens. M. S. Vats (1940) dated the sesame to about 3500 to 3050 B.C. but this was before the advent of radiocarbon dating, and the dates are probably much too early (see Bedigian 1985 and Bedigian and Harlan 1986 for reviews of the evidence and its archaeological context). The Indus Valley, however, is probably the area where the plant was first cultivated.

The Preclassical Near East

Mesopotamia. To date, no sesame seeds have been recovered from excavations in Mesopotamia.

Egypt: King Tutankhamen's Tomb. The discovery of sesame seed remains associated with the burial of King Tutankhamen (around 1350 B.C.) pushes back the date of the earliest record for sesame seeds in the Near East (Vartavan 1990). Some 30 boxes of plant remains from his tomb were stored at the Royal Botanical Gardens, Kew, England, where botanist Leonard Boodle worked on cataloging these specimens until his death in the 1930s.

Recently, Christian L. T. de Vartavan, a graduate student, had the good fortune of being allowed to go through these boxes to examine the contents. A letter to Bedigian from Mr. Vartavan, dated July 20, 1988, stated that there were 60 milliliters (ml) of sesame seeds, which were the main ingredients of one of the many containers in the tomb filled with food, drinks, ointments, perfumes, and oils. Vartavan viewed these seeds as "the most striking find" among the remains from the room designated by Howard Carter (who discovered the tomb in 1927) as the "Annexe" to Tutankhamen's tomb. These seeds from the Annexe represent the only find of sesame ever recorded from investigations of ancient Egypt, as well as the earliest sesame from the Near East and from Africa.

Armenia: Karmir Blur. Remains of sesame seeds dated between 900 and 600 B.C. have been found at this early Iron Age Urartian site on the outskirts of Yerevan. Four large jars containing carbonized sesame seed were excavated. Elaborate devices for the extraction of sesame oil indicate that the Urartians processed the seeds for oil (Piotrovskii 1950, 1952;

Kassabian 1957; Bedigian 1985; Bedigian and Harlan 1986).

Jordan: Deir Alla. Some 200 sesame seeds dating from about 800 B.C. were uncovered in Iron Age beds at this site in Jordan, according to R. Neef (1989).

Turkey. Sesame seeds were recovered at the Gordion ruins in Anatolia during the 1989 excavation of the "Destruction Level" (around 700 B.C.). The seeds are from a pure sample (about 250 ml) from within a pot that was found on the floor (along with other pots containing pure wheat, barley, and lentil). Massive burning and the collapse of the roof seems to have effectively sealed the contents on the floor (N. Miller, personal communication, 1990).

Early Literary Records

The Indian Subcontinent

The fact that the Sanskrit words for oil and for sesame seed are the same suggests that sesame may be one of the earliest oil-bearing plants to have been brought under cultivation (Sampson 1936: 221). One of these words, *tīla*, has been employed in religious ceremonies from very early times. It was regarded as holy (Hopkins 1968), and an offering of *tīla* seeds was considered effective in removing sins (Gupta 1971). Offerings of water and sesame were said to free an individual of all debts to his ancestors.

The Near East

Mesopotamia. Early Old Babylonian (OB) documents contain numerous references to sesame oil (Simmons 1978). Texts list the expenditure of sesame oil "for the inner bolt," "for the fire offering," "for the prince," "for the royal purification rite," "for the inner bolt on the day of Akitu," "for the sizkur divine name," "for the Elunum divine name," "for the regular offering," and "for anointing the banner." These are all special cultic applications that employed the oil to lubricate, soap, or fuel someone or something, almost certainly at springtime festivals (W. Doyle, personal communication, 1984).

The article for *ellu* in the *Assyrian Dictionary of the Oriental Institute, Chicago* (better known as the *Chicago Assyrian Dictionary* [CAD], ed. Oppenheim et al. 1958) provides as one definition: "[C]lean, pure in connection with oil, etc. . . . fine oil . . . sweet oil . . . pure sesame oil, sesame oil of the first [pressing]," used for anointing and making perfume. Another definition is: "[H]oly, sacred."

The *šamaššammū* article prepared for the CAD's "š" volume (Reiner ed. 1989: 301-7) mentions several texts that help to identify *šamaššammū* as sesame. The article itself contains many references to oil pressing, including one text (Keiser 1917) that specifies *šamaššammū pesûtu* (white *šamaššammū*). The

texts concerning white-seeded *šamaššammū* provide evidence of considerable importance in helping to distinguish flax from sesame because there are no flax cultivars with white seeds (J. Miller, personal communication, 1984).

Egypt. The earliest Egyptian textual reference to sesame seems to have been in the Tebtunis Papyri 3 (Part 2, No. 844) that dates from 256 B.C. (Lucas 1962: 336). H. Deines and H. Grapow (1959) indicated that sesame was used as a medicine. Pliny (1938: 15.7.31) wrote that a large amount of oil in Egypt was obtained from gingelly (*sesamum*).

The difficulty is to identify the ancient words for sesame. There is a striking linguistic resemblance of the Mesopotamian word *šamaššammū* to related Near Eastern terms, such as the Arabic word *simsim* and the name of a plant with edible seeds that is transcribed *smsmt* (Germer 1979). V. Loret (1892) regarded the Coptic name for sesame, *oke*, as Egyptian in origin. Another word from the hieroglyphs, *ake*, referred to a plant that produced oil and had seeds that were used medicinally. *Ake*, then, could be the Egyptian name for sesame (Loret 1892). But whether this later became the highest-quality oil, which was *nbb*, that one encounters from the nineteenth dynasty (1320 to 1200 B.C.) onward, remains a mystery. The assertion by some authors that *nbb* was *Ricinus* oil has been disputed on the grounds that castor oil is unpalatable (Keimer 1924) and also toxic.

Classical Greece

Sesame was cultivated extensively in the Greco-Roman world, but more for its edible seed than for its oil. The writings of Greek travelers and historians make it clear that sesame was well known in Mesopotamia by the time of the Iron Age, which began in about the eighteenth century B.C. About 300 years later, Herodotus (1928) observed that the only oil the Babylonians used was from sesame.

The cultivation of sesame in ancient Armenia was documented in the fifth century B.C. by Xenophon, who wrote: "In (western Armenia) . . . there was a scented unguent in abundance that they used instead of olive oil, made from pork fat, sesame seed, bitter almond and turpentine. There was a sweet oil also to be found, made of the same ingredients." Xenophon also placed sesame in two other parts of Anatolia. One was Cilicia - "[t]his plain produces sesame plentifully, also panic and millet and barley and wheat" (1901: I.ii.22) - and the other was "Calpe Haven in Asiatic Thrace," farther west. "Calpe lies exactly midway between Byzantium and Heracleia," has "good loamy soil . . . produces barley and wheat, pulses of all sorts, millet and sesame, figs in ample supply, numerous vines . . . indeed everything else except olives" (I.ii.22).

Imperial Rome

Cultivation requirements. Columella, in the first century A.D., reported accurately that “it [sesame] usually requires loamy soil, but it thrives no less well in rich sand or in mixed ground. . . . But I have seen this same seed sown in the months of June and July in districts of Cilicia and Syria, and harvested during autumn, when it was fully ripe” (1941: 2.10.18). He also wrote that “[i]n some districts [of Anatolia] such as Cilicia and Pamphylia, sesame is sown this month [late July to August]; but in the damp regions of Italy it can be done in the last part of the month of June” (11.2.56). This report would seem to indicate that, like the Babylonians, the Cilicians and Pamphylians grew sesame as a second crop after harvesting barley or another earlier crop.

Pliny, also writing in the first century A.D., said that gingelly [*Sesamum*] and Italian millets were summer grains and that gingelly came from India, where it was also used for making oil; and the color of the grain was white. He appears to have known sesame well. His advice for soaking the seeds prior to milling is reminiscent of the practice in Urartu: “Gingelly is to be steeped in warm water and spread out on a stone, and then rolled well and the stone then dipped in cold water so that the chaff may float to the top, and the chaff again spread out in the sun on a linen sheet, before being rolled again” (Pliny 1938: 18.22.98). In addition, E. L. Sturtevant (1972) stated that the Romans ground sesame seeds with cumin to make a pasty spread for bread.

Urartu (Armenia)

Based upon his participation in the excavation of an oil-extraction workshop at the Urartian site of Karmir Blur near Yerevan, Z. Kassabian (1957: 113) reconstructed Urartian techniques for sesame oil production as follows:

Sesame reserves were brought to the oil press. They were washed in a basin-shaped stone container, 79 cm in diameter, carved from a block of tufa. The basin joined a cylindrical pipe made of the same stone, that allowed waste liquid to drain out beyond the citadel. Sesame seeds were soaked to ease the removal of the tegument. After maceration and thorough pressing, the sesame was moved in a semi-moist condition to the oil press (workroom #2). Here they pounded the sesame using mortars and pestles.

Details about the plant remains and tools uncovered at Karmir Blur have been summarized by Bedigian and Harlan (1986: 146-7). The workrooms were furnished with fireplaces for parching the seed. Other finds included clay storage jars 1.5 meters tall, cakes of pressed sesame (the solid residue that remains after seeds are crushed for oil), and stone mortars, pestles, and graters.

Arabia

Language. Gingelly, a name for sesame that is still often used today in India and Europe, is derived from the Arabic *jujulân* (Dymock et al. 1893; Gove 1967). Spaniards say *ajonjoli*, the French *juggleone*, and present-day Arabic medicinal and botanical works employ both *al-jujulân* and *simsim*. The word *jujulân* was in use by the eighth century, as evidenced in a poem (quoted in *Lisan al-'Arab* [1981]) by Waddâh el Kubani-al Yamani (died 709) (Faroukh 1965): “The people laughed and said: ‘The poetry of Waddâh the Kinanite!’ My poetry is only salt mixed with *jujulân*.”

Clearly, *jujulân* had the meaning of tiny seeds, and sesame was a plant proverbial for its heavy production of tiny seeds (Charles Perry, personal communication, 1995). But *jujulân* is usually defined as “sesame before the seeds are removed,” that is, the capsule. Abu al Gauth said, “*Al-jujulân* is sesame in its hull [or peel], before it is harvested.” Charles Perry (personal communication, 1995), a specialist in Near Eastern languages, has discussed the relationships among the various terms. He wrote: “The Hebrew *shumshom* (mentioned in the Mishna but not in the Bible), Aramaic *shushma*, Armenian *shusham*, Turkish *susam*, Arabic *simsim*, Greek *sesamon*, and the rest go back to the word recorded in Sumerian (though whether it’s really of Sumerian origin or borrowed from Akkadian, the Semitic language spoken by the Sumerians’ neighbors, is a moot question; there were borrowings in both directions). This is the usual Arabic word for sesame. For instance, in the *Arabian Nights*, when Ali Baba says ‘open, sesame’ he says *iftab, ya simsim*.”

Symbolism in legend. The Sudan Department of Agriculture and Forests and Department of Economics and Trade bulletin on sesame (1938) opens with a concise version of “Ali Baba and the Forty Thieves”:

When the robbers had departed Ali went to the door of the cave and pronounced the magic words he had heard them use, “OPEN SESAME.” The door opened and he went inside and the door closed behind him. So astonished was he at the sight of the treasures in the cave and so absorbed in contemplation of them that when at last he desired to leave the cave he had quite forgotten the formula “OPEN SESAME.” In vain he cried aloud Open wheat, Open barley, Open maize, Open lentils; none of these availed and the door remained shut.

The significance of sesame in Arab culture is suggested by the fact that it was chosen as “a magical means of commanding access” (Arulrajan 1964). Once, sesame was thought to have mystical powers, and for some it still retains a magical quality. In fact, “open sesame” has become a common cliché that is

still used today. But the question "why sesame?" remains, and the answer might be that the high-quality sesame oil could have been thought to act magically to oil locks and open doors; in addition, sesame capsules do dehisce spontaneously (magically?) to release their seeds.

Persia

Symbolism in legend. Another example, this one of sesame employed as an omen, comparable to its usage in the magical incantation "Open Sesame," occurs in the *Iskandarnama* (Book of Alexander), one of the *Sbarafnama*, the "Book of Kings," completed A.D. 1010. Accompanied by an illustration of a miniature of Alexander the Great feeding sesame seed to birds (Titley 1979: 8, 10-11), the story

relates the parleying and battle between Alexander the Great and Darius. The Sultanate *Sbarafnama* in the British Library has a rare illustration of an episode in which Darius insulted Alexander by sending him a polo stick and ball and a bowl of sesame seed. Darius was angry because Alexander had not sent him gifts in the traditional manner and despatched a messenger to tell him so. Alexander equally angry replied that Darius had treasure enough already whereupon Darius sent him the polo stick and ball and a bowl of sesame seed saying that as Alexander behaved like a child he should have the playthings of a child. The sesame seed represented the countless soldiers in the great army Darius proposed to send against him. Alexander chose to interpret the gifts in another way and saw them as omens of victory. To him the polo ball represented the world (i.e. Darius' possessions) which Alexander would draw towards himself with the stick (i.e. by means of his army) as in polo. He threw the seed to birds which pecked every grain from the ground and he told Darius that it would be thus that his soldiers would wipe out the army of Darius. He then sent the messenger back to Darius with a bowl of mustard seed as a symbol of his own soldiers. The miniature graphically portrays this incident with a flock of hoopoes, parrots, pigeons, starlings and crows pecking the grain watched by Alexander and his retinue while the polo sticks and bowl of remaining seed are in the background (Titley 1979: 10-11).

China

According to the *Pen Ts'ao Kang Mu* (1596) by Li Shih-Chen, a classic ancient Chinese herbal and medical treatise, sesame was brought from the West by General Chang Ch'ien during the Han dynasty (second century B.C.), probably via the Silk Route. The Chinese name *tzub-ma* indicates introduction from overseas, and the Han dynasty was marked by expan-

sion that opened China to things foreign, including foreign foods (Yü 1977: 80). Y. S. Yü has written that because sesame appears three times in the text, it seems to have been particularly important. Moreover, he suggests that the "barbarian grain food" (*bu-fan*) enjoyed by Emperor Ling (A.D. 168-88) was most likely grain food cooked with the flavorful sesame.

Yü (1977: 81) quotes another source (Ch'i 1949: 294-5): "Under the Later Han, a great variety of noodle foods were cooked, including boiled noodles, steamed buns (modern *man-t'ou*), and baked cakes with sesame seeds." In T'ang times these were extremely popular foreign cakes, and a steamed variety containing sesame seeds was particularly well liked. In the capital city, these cakes were sold by foreign vendors - seemingly Iranians for the most part - on street corners (Hsiang 1957: 45-6, cited in Schafer 1977: 98). F. W. Mote listed sesame-oil noodles among a group of sacrificial food offerings made to ancestors during the Ming dynasty and suggested: "[W]e can assume that they reveal the tastes and food ideals of the former poor peasant family which now found itself the imperial family, and that the foods offered were also those actually eaten in the imperial household" (Mote 1977: 217).

Record of Modern Usage

India

Sesame is employed extensively in India, especially in the states of Gujarat and Tamil Nadu. Among other things, the seeds, with rice and honey, are used to prepare funeral cakes (*pindas*) that are offered to the ancestors in the Sraddh ceremony by the Sapindas (Dymock et al. 1893). *Tilanna*, sesame-rice balls formed in the shape of cows, are offered to relatives and friends of the deceased after a funeral. This ritual is enacted to say a proper farewell to the departed, and, as mentioned, the offering of sesame seeds is considered effective in removing sins (Gupta 1971). Indeed, *tilanjali* is a derived word meaning "to bid a final good-bye/to leave" (Chaturvedi and Tiwari 1970).

There are also many other uses. According to Dymock et al. (1893: 26-7): "On certain festivals six acts are performed with sesame seeds, as an expiatory ceremony of great efficacy, by which the Hindus hope to obtain delivery from sin, poverty, and other evils, and secure a place in Indra's heaven." These acts are *tilodvarti* ("bathing in water containing the seeds"), *tilasuayi* ("anointing the body with the pounded seeds"), *tilabomi* ("making a burnt offering of the seeds"), *tilaprada* ("offering the seeds to the dead"), *tilabhuj* ("eating the seeds"), and *tilavapi* ("throwing out the seeds"). In proverbial language, a grain of sesame signifies the least quantity of anything. Examples of this usage are *til chor so bajor chor* ("who steals a grain will steal a sack") and *til til ka bisab* ("to exact the uttermost farthing") (Dymock et al. 1893: 27).

S. M. Vaishampayan (personal communication, 1977) has noted that at the festival of "Makar Sankranti," which takes place on January 14 (when the sun enters the Zodiac sign "Makar," or Capricorn), the Hindus eat a composition of *gur* (brown sugar) and *til*, which they call *tilkut*. Among the Maharashtra Brahmins, elders offer sweetmeats made out of sesame seed and powdered sugar to youngsters, and bless them, saying: "*Tilghya, Gulghua* and *Gode bola*," which means "have sesame sweets and be sweet with everybody."

Donald Lawrence, Professor Emeritus of Botany at the University of Minnesota, who corresponded with this author about the uses of sesame in India, sent some beads from Pune, India (courtesy of Makarand S. Jawadekar), that are made of sesame seeds encrusted with sugar. Jawadekar, in a letter to Lawrence, dated February 9, 1986, wrote: "Newly wed brides wear, on January 14th, the necklace made out of these sugar-sesame beads - believed to bring good luck."

The Middle East and North Africa

This region shares common foods and similar preparation methods as a result of a historically steady exchange of ideas as well as trade goods. The peoples of Arabia, of the emirates in the Persian Gulf, of Armenia, Egypt, Greece, Iraq, Iran, Israel, Jordan, Lebanon, Somalia, Sudan, and Syria all prepare versions of *hummus* and *baba gannouj*, combining the flavors of chickpea and smoky eggplant with sesame paste, *tabini*, as well as sesame candies and biscuits.

The word for sesame in the Armenian dictionary (Yeran n.d.) is *shushmah*, not unlike the Sumerian word *šamaššammū* (confirmed by Charles Perry, personal communication, 1995). There are two versions of an Armenian coffee cake prepared with sesame. One is *pagharch*, a leavened bread in which the dough is folded with sweetened sesame paste.

Another version is a specialty item prepared just before Armenian Easter. A soft dough is allowed to rise to double its size, and a sesame mixture is brushed upon it. Then the dough is rolled like a slender jelly roll, which is finally coiled into a flat round circle and allowed to rise again before baking. Bread rings topped with sesame seed are sold everywhere in Turkey and take their name, *simit*, from *simsim*. In Morocco, sesame is known as *jinjelan*. It is used in Moroccan bread and desserts and, when toasted, the seeds are a popular garnish for chicken and lamb *tagines*.

In Sudan, sesame used for oil and seed constitutes one of the country's food staples (Bedigian and Harlan 1983). An enormous genetic diversity of traditional sesame cultivars in Sudan seems to correspond with the region's cultural diversity (Bedigian 1991). People are extremely fond of sesame as both a food and a metaphor for something small and precious. Sesame is used in the form of a solid cake blended with sugar, called *takneeya* in the Sudan; the name for

the sesame paste alone is *tabeena*. *Tabeena* is employed as a sauce, with fresh tomatoes and cucumbers, or with *tamia*, a fried chickpea patty to which sesame seeds can be added.

Egypt takes 90 percent of Sudan's exported sesame to make *balvab*. Rural growers in Sudan deliver their sesame seeds to oil presses called *assara*, powered by camels or engines (Sudan Department of Agriculture and Forests and Department of Economics and Trade 1938; Bedigian and Harlan 1986).

East Africa

Sesame, *nyim*, is one of the traditional crops of the Luo of western Kenya, and it is very likely that the crop arrived there when the first Luo migrated from the Sudan (Ogot 1967). Yet today, there is substantially less cultivation of sesame than there was at the beginning of the twentieth century. Most of the sesame for sale in western Kenya comes from Uganda, and such a decline in sesame cultivation is not a singular event. Similarly, finger millet, *cal*, which has historical significance for the Luo and was probably brought to Kenya by them during the earliest migrations from the Sudan, is now being replaced by maize, despite the fact that millet is more nutritious. Much of this has to do with British colonialism, which replaced sesame with corn oil or even less healthy hydrogenated shortenings.

When sesame is used in Kenya, it is pounded into a paste to be served with sweet potatoes or with local leafy greens. The paste can be diluted and used as a dip called *dengo*; it may also be added when cooking fish. A *winja*, combining black sesame seeds, finger millet, and dry beef cooked in sesame oil, is a special dish for a new son-in-law before his wedding. The dish is also used to appease spirits and to prevent lightning strikes - in short, to ward off harm.

China and the Far East

The Chinese considered sesame to be the best of the oilseeds, and their name for sesame oil (*bsiang yu*) means "fragrant oil." According to Bray (1984: 524-5): "Not only does it have all the qualities mentioned in the *Thien Kung Khai Wu*, the expressed cake is a protein-rich livestock food (sometimes eaten, fermented, by humans in India and Java), and in China it was recognized as an excellent fertilizer." B. Cost (1989) called sesame seed oil "a contender for the world's most seductively flavored oil, a seasoning fundamental to the cuisines of China and Japan. Like the Chinese sesame paste, the oil is pressed from roasted seeds. Cooking with it is a waste, since it loses its flavor over high heat and is expensive."

Regarding sesame seed paste, Cost (1989) has written that (unlike the Middle Eastern *tabini*) the Chinese paste is made from roasted, rather than raw, sesame seeds. As a flavorful peanut butter-like sauce, it is used mostly in sauces and dressings for cold dishes and salads, but occasionally in marinades. It is

also a constituent of *shichimi* ("seven spices"), consisting of the dried orange berry of an indigenous ash (related to the Sichuan peppercorn, and ground dried chillies), flakes of dried orange peel, white poppy seeds, sesame seeds, black hemp seeds, and bits of seaweed. This is a mixture that can be sprinkled over a variety of dishes.

Japanese sesame tools. Y. Kakehashi (personal communication, 1995) has indicated that, in Japan, fresh sesame seeds are toasted prior to adding them to a dish to enhance its flavor. The Japanese use a clay container, designed specifically for this purpose, called *goma iri* (*goma* is sesame). It is glazed inside and is approximately 10 centimeters in diameter and 20 centimeters in length, including the handle. The shape is hollow (like a rubber tire) except for its completely flat bottom; the sides fold inward toward the open center, preventing the seeds from escaping as they pop during the parching. The toasted seeds are collected by pouring them out of the conical funnel-shaped hollow handle. There is also a special ceramic mortar, *suribachi*, for grinding sesame seeds. It has corrugations inside that allow gentle crushing to take place.

Europe

Mary Tolford Wilson (1964) has noted the mercantilistic encouragement the British provided for cultivating sesame in their American colonies. The hope was that sesame oil might replace the olive oil they imported from European rivals, and thus sesame became one of those commodities whose colonial production the British encouraged. By contrast, the use of sesame was discouraged and even prohibited in nineteenth-century France in order to protect local rapeseed-growing interests, a policy that hints at the popularity of sesame at that time.

The United States

If the British contemplated the production of sesame oil in America, it may have been Africans who brought the seed to their attention. One name applied to sesame in the southern United States is *benne*. Lorenzo Turner (1969: 62) has compiled a comprehensive list of Africanisms in the Gullah dialect, in which *bene* is a feminine personal noun that means sesame in Bambara (French West Africa) and Wolof (Senegal and Gambia). Jessica Harris (personal communication, 1995), the culinary historian whose specialty is African-American foods, was astonished to learn this fact. She has not seen many examples of African-American words for foods in which the original language can be so specifically identified.

It is the case, however, that the contributions of African slaves to crop introductions in the Americas have rarely received sufficient recognition. In their free time, slaves tended their own gardens, where

they grew vegetables, including greens and "presumably African favorites such as okra, sorghum, black-eye peas, eggplant and benne seed" (Hess 1992: 8). These crops were very likely brought from Africa during the slave trade, along with yams, watermelons, and indigo.

That sesame was in colonial America by 1730 may be seen in a document in the *Records in the British Public Record Office Pertaining to South Carolina*. A letter from one Thomas Lowndes, dated August 26, 1730, reads:

My Lords, a planter in Carolina sent me some time ago a parcel of seed, desiring I would try it, and see of what use it would be, for if it turn'd to account South Carolina could with ease produce any quantity of it. By an experiment I found twenty one pounds weight of seed produced near nine pounds of good oyl, of which more than six pounds were cold drawn and the rest by fire. The name of the seed is Sesamum it grows in great abundance in Africa and Asia, and the inhabitants of those parts eat it, as well as use it for several other purposes. Pliny and many other good authors ancient and modern treat of this seed. It rejoyses in the Pine Barren Land (which is generally a light sandy soil). . . . This seed will make the Pine Barren Land of equal value with the rice land . . . and is for many purposes even preferable to oyl olive.

Some decades later, on May 25, 1774, the *Georgia Gazette* printed an early shipping record of sesame (page 2, column 2) aboard the ship *Savannah*, which carried 500 pounds of "benny" along with 20,000 pounds of sago, 200 gallons of soy, 200 pounds of vermicelli, 1,000 pounds of groundnuts, and a 10-gallon keg of sassafras blossom.

The most famous figure of America's early years to be interested in sesame was Thomas Jefferson, who earnestly recommended sesame growing to his friends (Wilson 1964). Jefferson's "Garden Book," which is the richest single source on the crop and records plantings of sesame grown at Monticello as long as the book was kept (to 1824), was edited and published by E. M. Betts in 1944. In it is a letter from William Few of New York to Jefferson, dated January 11, 1808, that was not transcribed by Betts but cited this way: "Tells Jefferson how to extract oil from Beni seeds and the history of its introduction to America." A facsimile of the original manuscript was provided to this author by the Manuscript Division, Library of Congress. Here is the text of the unpublished document:

When introduced into the southern states? We have no certain account of a tradition of its introduction. The Negroes are well acquainted with it and fond of cultivating it. It was probably introduced from the Slave Coast, but as we

know that it was cultivated up the Mediterranean in very early time, it is possible that we may have received it from that quarter. If the correct orthography of the name could be ascertained it might determine the question. We know that we received our rice from both of these sources and it may have been the same case with the benne. When it was introduced it can be difficult to ascertain as it has never interested us as an object of agriculture or curiosity. We only know that the oldest of our inhabitants remember it from their earliest days.

The Caribbean

Trinidad. *Bene* (sesame) is included in the lexicon of Wolof contributions to the language in Trinidad (Warner-Lewis 1991). Africans in nineteenth-century Trinidad had a ceremony of thanksgiving and ancestor intercession called the *saraka*, during which animal sacrifice and offerings of unsalted food were made. Ritual plant ingredients for the ceremony included kola nut (sent as a token to guests with their invitations), rum, stout and wine, *bene* (called *ziziwi*), black-eyed peas and rice, corn cooked with olive oil, and bread. Similarly, a Yoruba wedding in central Trinidad was characterized by a feast that included *bene* made into balls and sugar cake (Warner-Lewis 1991). Rice, black-eyed peas, potato, and grated, dried cassava also formed part of the menu.

Jamaica. That Africa was the source of sesame reaching Jamaica is reflected in a local name, acknowledged in a publication by J. Lunan as early as 1814. Sesame was called "*sesamum Africanum*," brought to Jamaica under the name of *zezegary*, and Lunan reported that "the first time I saw the plant, it was growing in a negro's plantation, who told me, they ground the seed between two stones, and eat it as they do corn. Their seed-vessels, [are] full of small white seeds, which the negroes call *soonga*, or *wolongo*. The oil that is drawn from it is called *sergilim* oil. The seed is often mixed and ground with coco, to make chocolate" (Lunan 1814: 252). Sesame was prized as a source of oil, with good reason:

[T]here are few which more deserve to be extensively cultivated into general domestic use, [to replace] that abominable rancid butter imported hither from Europe. Nothing but the grossest prejudice, in favor of old habits, can influence the inhabitants to persevere in the importation of that unwholesome, nauseous stuff, and to swallow it every day with their food, when they may supply themselves with so fine, nourishing and wholesome an oil, as the *sesamum*, for an ingredient in their pastry; nor are they less blamable, for continuing to import the olive oil, which is generally rancid before it

arrives and fitter for perukes [wigs] than salads (Long 1774: 809).

Dorothea Bedigian

Bibliography

- Anon. 1774. Records of the states of the U.S. *Georgia Gazette*, May 25. Savannah, Ga.
- Aristophanes. 1930. *The eleven comedies*, anon. trans. New York.
- Arulrajan, R., comp. 1964. *English-Tamil senior dictionary*. Madras.
- Bedigian, D. 1984. *Sesamum indicum* L.: Crop origin, diversity, chemistry and ethnobotany. Ph.D. dissertation, University of Illinois.
1985. Is *se-gis-i* sesame or flax? *Bulletin on Sumerian Agriculture* 2: 159-78.
1988. *Sesamum indicum* L. (Pedaliaceae): Ethnobotany in Sudan, crop diversity, lignans, origin, and related taxa. In *Modern systematic studies in African botany*, ed. P. Goldblatt, and P. P. Lowry, 315-21. St. Louis, Mo.
1991. Genetic diversity of traditional sesame cultivars and cultural diversity in Sudan. In *Biodiversity: Culture, conservation and ecodevelopment*, ed. M. L. Oldfield and J. B. Alcorn, 25-36. Boulder, Colo.
- Bedigian, D., and J. R. Harlan. 1983. Nuba agriculture and ethnobotany, with particular reference to sesame and sorghum. *Economic Botany* 37: 384-95.
1986. Evidence for cultivation of sesame in the ancient world. *Economic Botany* 40: 137-54.
- Bedigian, D., D. S. Seigler, and J. R. Harlan. 1985. Sesamin, sesamol and the origin of sesame. *Biochemical Systematics and Ecology* 13: 133-9.
- Bedigian, D., C. A. Smyth, and J. R. Harlan. 1986. Patterns of morphological variation in sesame. *Economic Botany* 40: 353-65.
- Betts, E. M. ed. 1944. *Thomas Jefferson's garden book, 1766-1824*. Philadelphia, Pa.
- Boland, A. R. de, G. B. Garner, and B. L. O'Dell. 1975. Identification and properties of phytate in cereal grains and oilseed products. *Journal of Agricultural Food Chemistry* 23: 1186-9.
- Bray, F. 1984. *Agriculture*, Vol. 27 of *Science and civilization in China*, ed. J. Needham. Cambridge and New York.
- Campbell, J. 1952. *The portable Arabian nights*. New York.
- Chakraborty, P. K., S. Maiti, and B. N. Chatterjee. 1984. Growth analysis and agronomic appraisal of sesamum. *Indian Journal of Agricultural Science* 54: 291-5.
- Chaturvedi, M., and B. N. Tiwari, eds. 1970. *Hindi-English dictionary*. Delhi.
- Chaudhry, S. N. 1988. Characterization and processing of raw and roasted sesame oil. M.S. thesis, Chapman University, Orange, Calif.
- Colquitt, H. R. 1933. *The Savannah cook book*. Charleston, S.C.
- Columella, L. J. M. 1941. *On agriculture*, trans. H. B. Ash. Cambridge, Mass.
- Cost, B. 1989. *Ginger east to west*. Berkeley, Calif.
- Crocomo, O. J., and L. C. Basso. 1974. Accumulation of putrescine and related amino acids in K-deficient *Sesamum*. *Phytochemistry* 13: 2659-965.
- Danilevich, A. F. 1977. Changes in content of free amino acids during roasting. *Kblebopekarnaya Konditerskaya Promsblennost* 2: 33-4.

- Danilevich, A. F., and A. Aboms. 1973. Changes in amine nitrogen level during fluidized bed searing of sesame seed kernels. *Izvestia Vyssbikh Uchebnykh Zavedenii Pishchevaya Tekhnologiya* 5: 76-9.
- David, E. 1970. *Spices, salt and aromatics in the English kitchen*. Harmondsworth, England.
- Dawkins, R. M. 1953. *Modern Greek folktales*. Oxford.
- Daxin, Z. 1994. The sesame oil mill. *Short Story International* 18: 4-43.
- Deines, H., and H. Grapow. 1959. *Wörterbuch der ägyptischen Drogennamen. Grundriss der Medizin der alten Ägypten*, Vol. 6. Berlin.
- Donner, H., and W. Roellig. 1964. *Kanaanaische and aramaische Inschriften*, Vol. 2. Wiesbaden.
- Dossin, G. 1933. *Lettres de la premiere dynastie Babylonienne*, Vol. 1. Paris.
- Dymock, W., C. J. H. Warden, and D. Hooper. 1893. *Pharmacographia Indica*, Vol. 3. London.
- Ebeling, E. 1923. Keilschrifttexte aus Assur religiösen Inhalts (KAR 376 r. 19). *Wissenschaftliche Veröffentlichung der Deutschen Orient-Gesellschaft* 34: ii-5.
- Ekpenyong, T. E., B. L. Fetuga, and V. A. Oyenuga. 1977. Fortification of maize flour based diets with blends of cashewnut meal, African locust bean meal and sesame oil meal. *Journal of the Science of Food and Agriculture* 28: 710-16.
- FAO (Food and Agriculture Organization of the United Nations). 1992. *Production yearbook 1991*, Vol. 45. Rome.
- Faroukh, O. 1965. *Tarikh al adab al Arab*. Beirut.
- Feigin, S. I. 1979. *Legal and administrative texts of the reign of Samsu-iluna*. New Haven, Conn.
- Fernández de Campoy, M. P. 1981. Development of food products utilizing the complementary protein sources of sesame seed (*Sesamum indicum*) together with either beans (*Phaseolus vulgaris*) or chickpea (*Cicer arietinum*). M.S. thesis, University of Arizona.
- Gallant, T. W. 1985. The agronomy, production, and utilization of sesame and linseed in the Graeco-Roman world. *Bulletin on Sumerian Agriculture* 2: 153-8.
- Gerhardt, M. I. 1963. *The art of story telling*. Leiden.
- Germer, R. 1979. Untersuchung über Arzneimittelpflanzen im alten Ägypten. Ph.D. thesis, University of Hamburg.
- Godfrey, S. A., B. J. Frances, and C. S. Kamara. 1976. Protein evaluation of cowpea and benniseed from Sierra Leone. *Tropical Science* 18: 147-54.
- Goetze, A. 1956. *Laws of Eshnunna*. New Haven, Conn.
- Gove, P. B., ed. 1967. *Webster's third new international dictionary of the English language*. Springfield, Mass.
- Gupta, S. M. 1971. *Plant myths and traditions in India*. Leiden.
- Herodotus. 1928. *History*, trans. G. Rawlinson, ed. Manuel Komroff. New York.
- Hess, K. 1992. *The Carolina rice kitchen: The African connection*. Columbia, S.C.
- Hilprecht, H. Unpublished OB tablet 1883: 25 and r. 15. Hilprecht Sammlung, Jena University.
- Hopf, M., and U. Willerding. 1989. Pflanzenreste. In *Bastan II*, ed. W. Kleiss, 263-318. Berlin.
- Hopkins, E. W. 1968. *Epic mythology*. Delhi.
- Hunt, A. S., and C. C. Edgar, eds. and trans. 1932. *Select papyri*, Vols. 1 and 2. Cambridge, Mass.
- Hussein, M. A., and M. M. Noaman. 1976. Comparison between white and red sesame seed and oil. *Olaj, Szappan, Kozmetika* 25: 5-7.
- Ibn Manzur Muhammed ibn Mukarram. 1981. *Lisan al-Arab*. Cairo.
- Ishikawa, S., and H. Tsuji. 1976. Seed coats of sesame removed with cellulase and hemicellulase. *Japan Kokai* 77 90, 647. Patent July 30, 1977. Appl. 76/3, 994. January 19, 1976.
- Johnson, L. A., T. M. Suleiman, and E. W. Lusas. 1979. Sesame protein: A review and prospectus. *Journal of the American Oil Chemists' Society* 56: 463-8.
- Johnson, R. H., and W. D. Raymond. 1964. Chemical composition of some tropical food plants. III: Sesame seed. *Tropical Science* 6: 173-9.
- Joshi, A. B. 1961. *Sesamum*. Hyderabad.
- Junior League of Charleston, comp. 1950. *Charleston receipts collected by the Junior League of Charleston*. Charleston, S.C.
- Kägi, M. K., and B. Wüthrich. 1993. Falafel burger anaphylaxis due to sesame seed allergy. *Annals of Allergy* 71: 127-9.
- Kassabian, Z. 1957. Production of vegetable oil in Urartu. In *Armenian. Izvestia Akademia Nauk Armenia SSR* 4: 107-16.
- Keimer, L. 1924. *Die Gartenpflanzen im alten Ägypten*, Vol. 1. Hamburg and Berlin.
- Keiser, C. E. 1917. *Letters and contracts from Erech written in the Neo-Babylonian period. Babylonian inscriptions in the collection of James B. Nies*. . . . New Haven, Conn.
- Laufer, B. 1919. *Sino-Iranica; Chinese contributions to the history of civilization in ancient Iran*, 288-96. In *Field Museum of Natural History Anthropological Series* 15. Chicago.
- Li Shih-Chen. [1596] 1893. *Pen ts'ao kang mu*. Shanghai.
- Little, M. 1991. Colonial policy and subsistence in Tanganyika 1925-1945. *Geographical Review* 81: 375-88.
- Long, E. 1774. *History of Jamaica*, Vol 3. London.
- Loret, V. 1892. *La flore pharaonique*. Rev. second edition. Paris.
- Lowndes, T. 1730. *Records in the British Public Record Office Pertaining to South Carolina* 14: 261-2.
- Lucas, A. 1962. *Ancient Egyptian materials and industries*. London.
- Lunan, J. 1814. *Hortus Jamaicensis*. Jamaica.
- Lyon, C. K. 1972. Sesame: Current knowledge of composition and use. *Journal of the American Oil Chemists' Society* 49: 245-9.
- Manley, C. H., P. P. Vallon, and R. E. Erickson. 1974. Some aroma components of roasted sesame seed. *Journal of Food Science* 39: 73-6.
- Mehra, K. L. 1967. History of sesame in India and its cultural significance. *Vishveshvaranand Indological Journal* 5: 93-107.
- Milne-Redhead, E., and P. Taylor. 1956. Description/label of *Sesamum indicum* specimen No. 9701, collected April 13, 1956, from the Songea district, Tanzania. Located in the East African Herbarium.
- Mirocha, C. J., S. V. Pathre, B. Schauerhamer, and C. M. Christensen. 1976. Natural occurrence of *Fusarium* toxins in feedstuff. *Applied Environmental Microbiology* 32: 553-6.
- Monacelli, R., E. Aiello, A. di Muccio, et al. 1976. Presence of aflatoxin B₁ in oleaginous seeds other than peanuts. *Rivista della Società Italiana di Scienza dell'Alimentazione* 5: 259-61.
- Monier-Williams, M. 1964. *Sanskrit-English dictionary*. New Delhi.
- Mookerji, R. K. [1912] 1962. *Indian shipping. A history of the sea-borne trade and maritime activity of the Indians from the earliest times*. Allahabad.
- Mote, F. W. 1977. Yüan and Ming. In *Food in Chinese culture*, ed. K. C. Chang, 193-258. New Haven, Conn.

- National Academy of Sciences. 1973. *Toxicants occurring naturally in foods*. Second edition. Washington, D.C.
- Nayar, N. M. 1976. Sesame. In *Evolution of crop plants*, ed. N. W. Simmonds, 231-3. London.
- Nayar, N. M., and K. L. Mehra. 1970. Sesame: Its uses, botany, cytogenetics and origin. *Economic Botany* 24: 20-31.
- Neef, R. 1989. Plants. In *Picking up the threads: A continuing review of excavations at Deir Allab, Jordan*, ed. G. van der Kooij and M. M. Ibrahim, 30-7. Leiden.
- Norman, J. 1995. *The complete book of spices*. New York.
- O'Dell, B. L., and A. de Boland. 1976. Complexation of phy-tate with proteins and cations in corn germ and oil seed meals. *Journal of Agricultural and Food Chemistry* 24: 804-8.
- Ogot, B. A. 1967. *History of the southern Luo*. Nairobi.
- Oppenheim, A. L., et al., eds. 1958. *The Assyrian dictionary of the Oriental Institute, Chicago*, "E." Chicago.
1962. *The Assyrian dictionary of the Oriental Institute, Chicago*, "š". Chicago.
- Parrot, F. 1845. *Journey to Ararat*. London.
- Pinches, T. G. 1899. *Cuneiform texts from Babylonian tablets in the British Museum* 8:8e:10. London.
- Piotrovskii, B. B. 1950. *Karmir Blur I*, in Russian. Yerevan.
1952. *Karmir Blur II*, in Russian. Yerevan.
- Pliny. 1938. *Natural history*, trans. H. Rackham. Cambridge, Mass.
- Prakash, O. 1961. *Food and drinks in ancient India*. Delhi.
- Purseglove, J. W. 1968. *Tropical crops: Dicotyledons*, Vol. 2. New York.
- Ratnagar, S. 1981. *Encounters: The westerly trade of the Harappan civilization*. Delhi.
- Reiner, E., ed. 1989. *The Assyrian dictionary of the Oriental Institute, Chicago*, Vol. 17, "š". Chicago.
- Romans, B. [1775] 1962. *A concise natural history of east and west Florida*. Gainesville, Fla.
- Rubenstein, L. 1950. Sensitivity to sesame seed and sesame oil. *New York State Journal of Medicine* 50: 343-4.
- Rutledge, S. [1847] 1979. *The Carolina bousewife*. Columbia, S.C.
- Salih, A. K.-E. M. 1993. Seed oils of *Sesamum indicum* L. and some wild relatives. A compositional study of the fatty acids, acyl lipids, sterols, tocopherols and lignans. Ph.D. thesis, Swedish University of Agricultural Sciences.
- Salunkhe, D. K., J. K. Chavan, R. N. Adsule, and S. S. Kadam. 1991. *World oilseeds: Chemistry, technology and utilization*. New York.
- Sampson, H.C. 1936. Cultivated crop plants of the British empire and the Anglo-Egyptian Sudan. *Kew Bulletin of Miscellaneous Information*, Additional Series 12: 220-5.
- Schafer, E. H. 1977. T'ang. In *Food in Chinese culture*, ed. K. C. Chang, 53-83. New Haven, Conn.
- Scheven, A. 1981. *Swabili proverbs: Nia zikiwa moja, kili-cho mbali buja*. Washington, D.C.
- Schoff, W. H., trans. 1912. *The Periplus of the Erythraean Sea*. London.
- Seegeler, C. J. P. 1989. *Sesamum orientale* L. (Pedaliaceae): Sesame's correct name. *Taxon* 38: 656-9.
- Sesame culture in the southern states. 1912. *Scientific American* 107: 276.
- Sharma, Y. K., A. D. Deodhar, and S. K. Gangrade. 1975. Chemical composition of sesame varieties of Madhya Pradesh, India. *Iran Journal of Agricultural Research* 3: 59-64.
- Simmons, S. D. 1978. *Early Old Babylonian documents*. Yale Oriental Series, Vol. 15. New Haven, Conn.
- Soden, Wolfram von. 1960. *Ellu. Akkadisches Handwörterbuch*, Vol. 1, 204-5. Wiesbaden.
- Soliman, M., S. Kinoshita, and T. Yamanishi. 1975. Aroma of roasted sesame seeds. *Agricultural and Biological Chemistry* 39: 973-7.
- Stol, M. 1985. Remarks on the cultivation of sesame and the extraction of its oil. *Bulletin on Sumerian Agriculture* 2: 119-26.
- Sturtevant, E. L. [1919] 1972. *Sturtevant's edible plants of the world*, reprint, ed. U. P. Hedrick. New York.
- Sudan Department of Agriculture and Forests and Department of Economics and Trade. 1938. *Sesame*. Sudan Government Department of Economics and Trade Bulletin No. 2. Khartoum.
- Tabekhia, M. M., and M. S. Mohammed. 1971. Effect of processing and cooking operations on thiamine, riboflavine and nicotinic acid content of some Egyptian national foods. II. Broad beans and sesame products. *Alexandria Journal of Agricultural Research* 19: 285-92.
- Theophrastus. 1916. *Enquiry into plants*, trans. A. Hort. Cambridge, Mass.
- Tinay, A. H. El, A. H. Khat-tab, and M. O. Khidir. 1976. Protein and oil components of sesame seed. *Journal of the American Oil Chemists' Society* 53: 648-53.
- Titley, N. M. 1979. *Plants and gardens in Persian, Mughal and Turkish art*. London.
- Tothill, J. D., ed. 1948. *Agriculture in the Sudan*. London.
- Turner, Lorenzo. 1969. *Africanisms in the Gullah dialect*. New York.
- Van Beek, G. W. 1969. *Hajar bin Humeid*. Baltimore, Md.
- Vartavan, C. de. 1990. Contaminated plant-foods from the tomb of Tutankhamun: A new interpretive system. *Journal of Archaeological Science* 17: 473-94.
- Vats, M. S. 1940. *Excavations at Harappa*. Delhi.
- Villegas, A. M., A. González, and R. Calderón. 1968. Microbiological and enzymatic evaluation of sesame protein. *Cereal Chemistry* 45: 379-85.
- Vishnu-Mittre. 1977. The changing economy in ancient India. In *Origins of agriculture*, ed. C. A. Reed, 569-88. The Hague.
- Wankhede, D. J., and R. N. Tharanathan. 1976. Sesame carbohydrates. *Journal of Agricultural and Food Chemistry* 24: 655-9.
- Warner-Lewis, M. 1991. *Guinea's other suns: The African dynamic in Trinidad culture*. Dover, Mass.
- Weiss, E. A. 1971. *Castor, sesame and safflower*. New York.
- Wilson, Mary Tolford. 1964. Peaceful integration: The owner's adoption of his slaves' food. *Journal of Negro History* 49: 116-27.
- Wolfert, P. 1973. *Couscous and other good food from Morocco*. New York.
- Xenophon. 1901. *March of the ten thousand* (Anabasis), trans. H. G. Dakyns. London.
- Yen, G.-C., and S.-L. Shyu. 1989. Oxidative stability of sesame oil prepared from sesame seed with different roasting temperatures. *Food Chemistry* 31: 215-24.
- Yeran, E. A. n.d. *English-Armenian pocket dictionary or pocket companion*. Third edition. Boston, Mass.
- Yermanos, D. M., S. Hemstreet, W. Saleeb, and C. K. Huszar. 1972. Oil content and composition in the world collection of sesame introductions. *Journal of the American Oil Chemists' Society* 49: 20-3.
- Yü, Y. S. 1977. Han. In *Food in Chinese culture*, ed. K. C. Chang, 85-140. New Haven, Conn.
- Zohary, D., and M. Hopf. 1994. *Domestication of plants in the Old World*. Second edition. Oxford.

II.E.5 Soybean

Description

The soybean plant belongs to the legume family (*Leguminosae*), the second largest family of flowering plants, with more than 14,000 species. Identifiable characteristics include a fruit located within a pod that dehisces along a seam from top to bottom. In the case of a soybean, the split takes place when the plant has matured and died to yield 2 to 4 seeds per pod that are easily removed without damage or loss.

The soybean's size, shape, and color is determined by the variety. Soybeans range in size from small (1 centimeter [cm]) to large (3.5 cm), can be flattened or oblong, and are colored yellow, green, brown, or black. They are approximately 8 percent hull, 90 percent cotyledon, and 2 percent hypocotyl (Wolf and Cowan 1971).

Terminology

The terms "soy and "soya" are said to have derived from the Japanese word *shoyu* (or *sho-yu*) that designates a sauce made from salted beans. But the Japanese word may well have been inspired by the ancient Chinese name for the bean, which was *sou*. In Chinese, the word for soy sauce is *jiangyou* (or *chiang-yiu*). C. V. Piper and W. J. Morse (1923) have recorded more than 50 names for the soybean or its sauce in East Asia. In English the bean has been called soya bean, soya, soy, Chinese pea, Japanese pea, and Manchurian bean, to provide just a few of its appellations. For the purposes of this chapter, soya is used synonymously with the soybean and its many products.

Early History

Present-day soybean varieties (*Glycine max*), of which there are more than 20,000, can be traced to the wild soybean plant *Glycine soja* that grew in abundance in northeast China and Manchuria (Hymowitz and Newell 1981). Legends abound concerning the discovery and domestication of this food plant that today is the most widely used in the world (Toussaint-Samat 1993: 51). Around 2700 B.C., the legendary Chinese emperor Shen Nung is said to have ordered plants to be classified in terms of both food and medicinal value, and soybeans were among the five principal and sacred crops (Shih 1959). This dating squares nicely with the judgment of modern authorities on Asian plants that soybeans have been cultivated for at least 4,500 years (Herklots 1972). But there are other sources that indicate that the domesticated soybean (*G. max*) was introduced to China only around 1000 B.C. perhaps from the Jung people who lived in the northeast (Trager 1995).

The court poems of the *Book of Odes*, sixth century B.C., also indicate that the wild soybean came from northern China and that its cultivation began around the fifteenth century B.C. Confucius, who died in 479 B.C., left behind writings that mentioned at least 44 food plants used during Chou times; they included soybeans. But they do not seem to have been very popular in ancient times. Soybeans were said to cause flatulence and were viewed mostly as a food for the poor during years of bad harvests. Nonetheless, soybeans were recorded in the first century B.C. as one of the nine staples upon which the people of China depended, and certainly there were enough people. The first official census conducted in Han China at about that time counted 60 million people, and even if such a number seems implausibly high – especially in light of a census taken in A.D. 280 that showed only 16 million – it still suggests that Chinese agricultural policies were remarkably effective, both in feeding large numbers of people and, one suspects, in encouraging the growth of large numbers of people (Chang 1977: 71). The famine in China in the year A.D. 194 may have been the result of too many mouths to feed and thus responsible, at least partly, for the discrepancy in the two censuses. But in addition, famine forced the price of millet to skyrocket in relation to soybeans, resulting in an increased consumption of the latter – often in the form of bean *conjee* or gruel (Flannery 1969).

Soybean



Early Dissemination

Because the wild soybean was sensitive as to the amount of daylight it required, and because the length of growing seasons varied from region to region, domestication involved much experimental planting and breeding to match different varieties with different areas. That this was done so successfully is a tribute to ancient Chinese farmers who, as noted, were doubtless impelled by an ever-increasing need to feed larger and larger populations of humans and animals. Soybeans ground into meal and then compressed into cakes became food for travelers and soldiers on the move who, in turn, widened knowledge of the plant.

Buddhist priests, however, were perhaps as instrumental as anyone in the domestication of the soybean and absolutely vital to its dissemination (Yong and Wood 1974). As vegetarians, they were always interested in new foods and drinks (such as tea, which they also nurtured to an early success in China). In their monasteries, they experimented with soybean cultivation and usage and found flour, milk, curd, and sauce made from soy all welcome additions to their regimes. As missionaries, they carried the soybean wherever they went, and in the sixth century A.D., they introduced it to Japan from Korea, which they had reached in the first century. Buddhism merged with the native Shinto religion, and the plant quickly became a staple in the Japanese diet.

Not only missionaries but also soldiers, merchants, and travelers helped introduce soybeans to Asian countries. The northern half of Vietnam had soybean food products as early as 200 B.C. During the sixth through the tenth centuries A.D., Thailand received soybeans from southwest China, and India was exposed to them during the twelfth century by traders from Pakistan.

Recent History and Dissemination

The Portuguese began trading in East Asia during the sixteenth century, as did the Spanish and later the Dutch. Yet the soybean was not known in Europe until the end of the seventeenth century when Engelbert Kaempfer published his *Geschichte und Beschreibung von Japan*, an account of his visit to that country during the years 1692–4 as a guest of the Dutch East India Company. He wrote of the bean that the Japanese prized and used in so many different ways, and in 1712, he attempted, not very successfully as it turned out, to introduce this miracle plant to Europe. Its products simply did not fit into the various cuisines of the continent, which, in any event, were only then in the process of fully utilizing the relatively new American plants, such as maize and potatoes.

The botanists, however, were thrilled to have a new plant to study and classify, and Carolus Linnaeus, who described the soybean, gave it the name *Glycine max*. *Glycine* is the Greek word for “sweet,” and “max”

presumably refers to the large nodules on the root system, although other sources suggest that the word *max* is actually the result of a Portuguese transcription of the Persian name for the plant (Toussaint-Samat 1992: 52).

Because of scientific interest, the soybean was shuttled about the Continent during the eighteenth century for experimental purposes. In 1765, soybean seeds reached the American colonies with a sailor named Samuel Bowen, who was serving aboard an East India Company ship that had just visited China. Bowen did not return to the sea but instead acquired land in Savannah, Georgia, where he planted soybeans and processed his first crop into Chinese vetch, soy sauce, and a starchy substance incorrectly called sago. In North America as in Europe, however, soybean products did not go well with the various cuisines, and the bean remained little more than a curiosity until the twentieth century, despite efforts to reintroduce it.

By the mid-nineteenth century, the soybean was being rapidly disseminated around the globe as trade, imperialism, clipper ships, and then steamships all joined to knit the world more closely together. The expedition of Commodore Matthew Perry that opened Japan to trade in 1853–4 returned to the United States with the “Japan pea” – actually 2 soybean varieties that were subsequently distributed by the U.S. Commissioner of Patents to farmers for planting. But lacking knowledge of and experience with the plant, the recipients were apparently not successful in its cultivation.

During the American Civil War, when shipping was disrupted, soybeans were frequently substituted for coffee, especially by soldiers of the Union army (Crane 1933: 270). Interest also arose in soybean cultivation as a forage plant, and the Patent Office and the new Department of Agriculture (USDA) encouraged experimental planting. The USDA's role in promoting agricultural research, regulating the industry, and serving as an information generator for farmers proved invaluable to all farmers and certainly to those growing soybeans for the first time (Arntzen and Ritter 1994: 466).

There were 2 stages in processing the soybean plant as a forage crop or hay. The first was to cut the plants just before the leaves turned yellow but after the pods, containing semiripened seeds, had formed, thereby increasing the plants' protein value. In the second stage, the plants were windrowed and left to dry for a day or two, after which the windrows were raked into bunches and dried for another three or four days. Lastly, the bunches were stacked in barns right side up. The average yield of hay using this method was approximately 4 tons an acre, with a protein content of close to 11 percent.

In harvesting the soybean plant for silage, the process called for cutting the plants earlier – when their seedpods contained premature green seeds.

These were made into bundles of about 25 pounds each, then stored in barns until needed. This method traded some protein content for less leaf loss during harvesting.

Soybean plants, processed by either of these methods, lowered the cost of feeding livestock by replacing the more expensive grass hay and corn. Even a combination of soybean hay or silage and traditional feeds resulted in considerably reduced feed costs and supplied more protein than hay or corn alone could deliver.

At the turn of the twentieth century, the population of the United States was swelling with immigrants, and significant technological advances were spinning out of an ever-accelerating industrial process. An increased demand for food spurred soybean cultivation, and processing was facilitated by electric motors to power grinding equipment that made soybean meal more quickly and efficiently than ever before. Mechanized farm implements encouraged the planting of still more land in soybeans, while lowering the costs of harvesting.

The result was that soybeans produced in the United States became competitive with those grown in East Asia, despite the Asian application of very cheap labor. By 1911, the U.S. industry not only processed soybean meal into cakes for livestock feed but also began to press the beans into oil, as China was already doing. Indeed, the high oil content of the soybean (about 20 percent) was arousing substantial commercial interest.

Previously, a shortage of soybeans in the United States and the predominance of cottonseed oil (then called Wesson Oil after David Wesson, who, in 1899, developed a method for clarifying it) had retarded the development of soybean oil. But in 1915, cottonseed oil became scarce because the U.S. cotton was infested with boll weevils, and this in turn led to the processing of soy oil for human consumption. Cottonseed-processing plants quickly became soybean-processing plants, because the presses and other equipment worked equally well with soybeans. Moreover, a new method was discovered for extracting the oil by first grinding the beans, then soaking them in a solution of benzol, naphtha, and ether, which for every bushel (60 lbs.) of soybeans yielded 10.5 pounds of oil and 48 pounds of meal. Thus, soybean oil was efficiently produced at a time when World War I was creating more demand for oils.

None of these lessons in supply and demand were lost on southern farmers, who began to plant soybeans on land barren because of the boll weevil. The USDA also encouraged soybean cultivation in the states of the Midwest. In 1922, a soybean-processing plant was opened in Decatur, Illinois. To ensure a steady soybean supply, the "Peoria Plan" was developed to guarantee Illinois farmers a base price of \$1.35 a bushel (Smith and Circle 1972). In addition, farmers throughout the nation were given an inducement

to grow soybeans with the passage, in 1923, of the Smoot-Hawley Tariff, which placed import duties of 2 cents per pound on soybeans, 3.5 cents per pound on soybean oil, and \$6.00 per ton on soybean meal.

Soybeans became the nation's "Cinderella" crop in the 1920s. Demand was high for soybean cakes, which continued to provide farmers with a high-protein, low-cost animal feed. But it was demand for soybean oil that stimulated still more production; soybeans yielded oil valued at 20 cents per pound, or \$400.00 per ton, as opposed to meal worth only \$20.00 a ton. Research supported by the processing plants helped plant breeders develop new soybean varieties with higher oil contents.

During the 1920s, this oil went into numerous industrial products, among them soaps, paints and varnishes, linoleum, ink, glycerin, explosives, celluloid, and a substitute for rubber. Moreover, the low cost of soybean production stimulated research to discover still more industrial uses. Yet, soybeans remained an underused food resource because of their relatively high saturated fat content, which made the oil solidify, as well as their high percentage of bad-tasting linolenic acid.

It was during the 1930s that research on soybean oil refinement, flavor reversion, and especially hydrogenation ultimately resolved these problems and opened the way for soybeans to be employed in food products (Toussaint-Samat 1993). Increasingly, the oil found its way into shortenings, margarine, salad dressings, and, of course, cooking oils. In fact, during the Great Depression years, soybean oil was well on its way to becoming the most important food oil in the United States, a status it achieved after World War II and never relinquished.

It is interesting to note that during that war, the only survival rations issued to Japanese soldiers had consisted of bags of soy flour, perhaps illustrative of the fact that despite the growing use of soybeans in the United States, they were still an Asian resource. Following the war, soybeans became the world's most important crop, not because of the Asian influence but because of productivity. Initially, much of the post-war surge of U.S. interest was due to the ability of soybeans to regenerate the soil when planted in rotation with corn (Pepper 1994: 193). During the war, soy margarine had replaced butter on most tables, a use that continued after 1945.

Making a transition to soybeans from a previous concentration on corn, wheat, cotton, or tobacco was scarcely a hardship for farmers because surpluses of these latter crops existed, and because with the passage of the Agriculture Adjustment Act in 1933, restrictions had been placed on the amount of acreage that could be devoted to them. Meanwhile, the government was promoting soybean products by, among other things, organizing the Regional Soybean Industrial Products Laboratory in 1938.

And once again research found industrial applications for soybeans – this time for using soy protein in the paper industry and for making plastics from oil-cake residue in the automobile industry. At the same time, world food requirements in the 1950s increased demand for soy protein for both humans and domesticated animals. During this decade, American farmers responded by producing an annual average of 300 million bushels of soybeans for industrial and food use. In the process, the United States, which prior to World War II had been the world's biggest importer of fats, became its greatest exporter, accounting for fully 75 percent of the world's soybean crop.

Because of plant breeding, the soybean – once a subtropical plant – moved north as far as 52 degrees, and soybean fields became familiar sights from Minnesota to the Deep South. Soybean meal processed in the northern states was carried by rail to barges on the Mississippi River that transported it to New Orleans for shipment to world markets (Forrestal 1982).

In addition to the ease of bulk transportation, soybeans lend themselves to handling in many other ways. They can withstand long storage and shipment over long distances; they are easily harvested in an entirely mechanical procedure; and a growing season of only 15 weeks makes it easy to adjust production to world market demand. In short, as one food author has pointed out, supply can be virtually guaranteed, and if there is overproduction, there is no need to destroy the surplus (Toussaint-Samat 1993).

The people of Far Eastern countries receive, on average, about 12 percent of their protein requirements from soybean products. But despite the high protein content of the soybean and the high quality of that protein, attempts to introduce soya to many poor regions of the world, such as India, Africa, and Latin America, have historically met with little success, with southern South America a notable exception. For example, in Mexico in the 1940s the National Indian Institute handed out soybeans to the impoverished Otomi Indians living in the Mezquital valley. Although the legume flourished in the arid soil, the experiment failed when Otomi women were unable to hand-grind the soybeans to make a decent tortilla (Granberg 1970).

International efforts, however, continue to increase consumption, particularly in the soybean fortification of cereals on which many people rely. All of these cereals – rice, maize, barley, wheat, and rye – yield a protein that is incomplete, meaning that it does not contain all of the essential amino acids. Yet what they lack, soybeans contain, and soybeans are also high in the B vitamins, along with vitamin E, phosphorous, and iron. Thus, in cases where soya is used to supplement other cereals, as in Golden Elbow or Vitalia macaroni, Kupagani biscuits, ProNutro cereal, and Cerealina, the consumer receives a whole protein equal to that provided by meat, fish, and dairy products (Fortified foods 1970).

Processing, Preparation, and Products

The oldest methods of preparing green, immature soybeans for consumption was by roasting or by soaking, grinding, and cooking (Toussaint-Samat 1993). Mature soybeans were processed in much the same manner but with a longer cooking time.

Soybeans, paradoxically, have a very bland taste, which probably inspired the development of flavorful fermented soy products. Such soy processing began in China during the Chou dynasty (1122–246 B.C.) and subsequently spread to other areas of the Far East.

One of the first of these products, and the best known to Westerners, is *shoyu* or soy sauce – a dark brown liquid used extensively in Chinese food preparation – that is obtained by fermenting a combination of soybeans and wheat. The original process entailed first boiling soybeans with *koji* (*Aspergillus oryzae*) which, in this case, was a mold skimmed from the surface of cooked wheat that had cooled and fermented. Soy sauce has a salty taste and a subtle, but tantalizing, aroma that goes well with rice dishes.

Another product is soy paste, produced in a process similar to that which yields soy sauce, except that the *koji* is derived from barley or rice. Soy paste is also Chinese in origin and evolved from a paste of fish puree called *jiang* that was used before soya became popular. By the time of the Han dynasty in the third century, soy paste had become the important ingredient in *jiang*. Two or three centuries later, the Koreans were producing soy paste and the Japanese got the recipe from them. Today there are numerous such pastes, called *miso*, in everyday use in Japan. The color can vary, depending on the soybean-to-rice or barley ratio. The greater the amount of rice used, the lighter (and sweeter) the product. In the past, soy paste was stored in earthen jars for up to a year, which was said to improve the flavor.

Tempeh kedlee is a third soy product primarily confined to Indonesia, where more than half of the soybeans produced are devoted to it. It has a fine flavor, but not the longevity of other soy products. Also Indonesian in origin is *ontjom*, made by combining the residue of soybean milk with peanuts and allowing the mixture to ferment. The result is a kind of soy sauce with a nutlike flavor.

Still another product that originated in Indonesia (and Thailand as well) is *tao tjo*, made by combining cooked soybeans and roasted rice flour with *koji*. After about four days, a fungus covers the solution, whereupon the mixture is dried in the sun, then soaked in brine. A few days later, sugar and rice yeast are added and the sun, once again, is employed to dry the combined beans and flour. The end product is a sweet soy sauce to accompany vegetable, meat, or fish dishes.

A final Indonesian contribution is *ketjab*, in which bacteria are combined with cooked black beans, fermented, and then placed in salt brine for a week or

so. The beans are drained and the residue is cooked several times. The extract that remains is sweetened with sugar and permitted to evaporate further.

A sixth soy product is *natto*, a gray-colored liquid with a strong, musty flavor, produced in Japan by wrapping cooked soybeans in rice straw and allowing the whole to ferment. The taste is something of an acquired one for Westerners and varies depending on locale.

Buddhist monks first developed *hamanatto* in their monasteries. This is still another fermented soy product, obtained in an elaborate procedure employed by the monks that began with soaking whole soybeans for up to 4 hours, then steaming them for 10 hours. Afterward, the beans were covered with *kofu* and left to ferment for 20 hours, during which time a green mold developed to cover the beans. Next, they were dried in the sun to lower their moisture content, then placed in wooden buckets with strips of ginger on the bottom. After a year, the soybeans were once more dried in sun – this time until they turned black. The final product was a soybean sauce with a very sweet flavor and a pleasant aroma.

The major Korean contribution to soy products of the world is *kochu chang*, made by mashing boiled soybeans and beginning their fermentation with the addition of a bit of a previous batch, much like starter dough. This mixture was then placed in a sack and hung to dry in the sun for a couple of months, after which time the fermented mash was pulverized and mixed with salt and water. The final step was to put the mixture in an earthen jar to age for three months or so. After chilli peppers reached the East in the sixteenth century, they were added to the mash for additional flavor.

In addition to these sauces, soybeans have been processed by the Chinese and others of the Far East into bean curd, soybean milk, and bean bran. Soybean curd or *tou-fu* is made by wet-grinding the beans into a thin mixture that is strained, then coagulated with gypsum, which causes the proteins to precipitate. The mixture is strained a last time to become bean curd and soy milk, respectively. Bean curd thus provides the Asians with flour and milk, both very important foods in everyday life. The curd itself is often served with rice, meats, vegetables, and fish. It is also added to soups, mashed for making breads and cakes, and deep-fried.

Soybean milk – a vegetable milk – is obviously of considerable importance in a region of the world where little cattle raising takes place and where the human population seems to be uniformly lactose intolerant. In addition to serving as a beverage in its own right, soy milk is processed as a soft drink (Trager 1995).

The sprouts of soybean and mung beans (*Phaseolus aureus*) serve as an instant vegetable high in vitamin C and are also blanched and processed into cello-

phane noodles. Soybean sprouts are traditionally eaten on the Chinese New Year. Finally, young soybean pods are eaten boiled or roasted as a snack with soy sauce or sesame oil.

Nutrition

East Asians have long been dependent on soybeans for corrections of nutritional deficiencies. Soy sauces and other typical soy products are all concentrates of the B vitamins and contain significant amounts of the minerals calcium, iron, magnesium, phosphorous, and zinc. In addition, as we have already noted, they yield a very high quality protein, as well as important lipids, carbohydrates, and fiber.

Soy protein is also important nutritionally because of the ways in which it is employed in other food products for purposes of emulsion formation, promotion of fat and water absorption, and texture and color control. Foods benefiting from soy protein include bologna, frankfurters, breads, soups, cakes, pasta, gravies, whipped toppings, and simulated meats, vegetables, and fruits (Wolf and Cowan 1971: 52).

Soy flour usage, however, has not lived up to earlier expectations. The flour does have a nutty taste and very fragrant aroma; sweetened with sugar, it is baked into breads, muffins, cakes, cookies, and the like (Piper and Morse 1923). But unfortunately, it has a limited shelf life, which discourages production in large quantities.

It would not do to end this section on the nutritional value of soybeans without noting that they also contain various potentially toxic substances that can inhibit growth, reduce the absorption of fats, enlarge the pancreas, and decrease the energy yield of the diet (Norman 1978: 227). Among these antinutritional agents are trypsin, phytic acid, hemagglutinin, saponin, and phenolic constituents. Fortunately, the process of cooking eliminates the toxicity of soybeans.

Thomas Sorosiak

Bibliography

- Adams, Sean. 1996. Sorting look-alike soybeans. *Agricultural Research* 44: 12-13.
- Adolph, William H. 1946. Nutrition research in China. *Journal of the American Dietetic Association* 22: 964-70.
- Anderson, Eugene N., Jr. 1977. Modern China: South. In *Food in Chinese culture*, ed. K. C. Chang, 70-80. New Haven, Conn.
- Arntzen, Charles J., and Ellen M. Ritter, eds. 1994. *Encyclopedia of agricultural sciences*. San Diego, Calif.
- Beuchat, Larry R. 1994. Fermented soybean foods. *Food Technology* 38: 64-70.
- Beuerlein, Jim, T. T. Van Toai, A. F. Schmitthenner, and S. K. St. Martin. 1987. *The soybean in Ohio*. Ohio Cooperative Extension Service. Ohio State University, Columbus.
- Borst, H. L., and L. E. Thatcher. 1931. *Life history and com-*

- position of the soybean plant. Ohio Agricultural Experimental Station. Wooster, Ohio.
- Carle, Julie. 1996. Research vital to soybean's future. *Sentinel-Tribune* (Bowling Green, Ohio). March 22, Farm edition, p. 11.
- Chang, K. C. 1977. *Food in Chinese culture*. New Haven, Conn.
- Crane, Helen R. 1933. The story of the soya. *Scientific American* 149: 270-2.
- Darwin, Charles. 1868. *The variations of animals and plants under domestication*. London.
- Dies, Edward J. 1943. *Soybeans: Gold from the soil*. New York.
- Erdman, John W., and E. J. Fordyce. 1989. Soy products and the human diet. *The American Journal of Clinical Nutrition* 49: 725-37.
- Flannery, K. V. 1969. Origins and ecological effects of early domestication in Iran and the Near East. In *The domestication and exploitation of plants and animals*, ed. P. J. Ucko and G. W. Dimbleby, 10-112. London.
- Forrestal, Dan J. 1982. *The kernel and the bean*. New York.
- Fortified foods: The next revolution. 1970. *Chemical Engineering News* 48: 35-43.
- Gould, S. J. 1980. *The panda's thumb: More reflections in natural history*. New York.
- Granberg, Wilbur J. 1970. *The Otomi Indians of Mexico*. New York.
- Hector, J. M. Introduction to the botany of field crops. *Central News Agency Ltd* 2: 697-708.
- Herklots, G. A. C. 1972. *Vegetables in Southeast Asia*. London.
- Hesseltine, C. W., and H. L. Wang. 1967. Traditional fermented foods. *Biotechnology and Bioengineering* 9: 275-88.
- Ho, Ping-ti. 1975. *The cradle of the East: An inquiry into the indigenous origins of techniques and ideas of neolithic and early historic China, 5000-1000 B.C.* Chicago.
- Hymowitz, Theodore. 1987. Introduction of the soybean to Illinois. *Economic Botany* 41: 28-32.
- Hymowitz, Theodore, and J. R. Harlan. 1983. Introduction of soybean to North America by Samuel Bowen in 1765. *Economic Botany* 37: 371-9.
- Hymowitz, Theodore, and C. A. Newell. 1981. Taxonomy of the genus *Glycine*; domestication and uses of soybeans. *Economic Botany* 35: 272-88.
- Liener, Irvine E., and M. L. Kakade. 1980. *Protease inhibitors*. New York.
- Mermelstein, Neil H. 1989. Seeds of change. *Food Technology* 46: 86-9.
- Norman, Geoffrey A. 1978. *Soybean physiology, agronomy, and utilization*. New York.
- Pepper, Gary E. 1994. Soybeans. In *Encyclopedia of agricultural science*, Vol 4: 193-202. New York.
- Piper, Charles V., and W. J. Morse. 1923. *The soybean*. New York.
- Rindos, David. 1984. *The origin of agriculture*. New York.
- Shih, Sheng-Han. 1959. "Fan Sheng-Chih Shu": An agriculturalist's book of China written by Fan Sheng-Chih in the first century B.C. Peking.
- Shurtleff, William, and A. Aoyagi. 1986. *Thesaurus for soya*. Second edition. Lafayette, Calif.
- Simoons, Frederick J. 1990. *Food in China*. Boca Raton, Fla.
- Smith, A. K. 1961. *Oriental methods of using soybeans as food, with special attention to fermented products*. U.S. Department of Agriculture. Washington, D.C.
- Smith, Allan K., and S. J. Circle. 1972. *Soybeans: Chemistry and technology*. Westport, Conn.
- Soskin, Anthony B. 1988. *Non-traditional agriculture and economic development: The Brazilian soybean expansion 1964-1982*. New York.
- Tanner, J. W., and D. J. Hume. 1976. *World soybean research*. Danville, Ill.
- Toussaint-Samat, Maguelonne. 1992. *A history of food*, trans. Anthea Bell. Cambridge, Mass.
- Trager, James. 1995. *The food chronology*. New York.
- Van Sanford, D. A. 1993. Selection index based on genetic correlations among environments. *Crop Science* 33: 1244-8.
- Wagner, C. K., and M. B. McDonald. 1981. *Identification of soybean cultivars using rapid laboratory techniques*. Ohio Agricultural Research and Development Center. Wooster, Ohio.
- Weber, C. R. 1966. Soybeans. *Agronomy Journal* 58: 43-6.
- Whigham, D. K. 1976. Expanding the use of soybeans. *Journal of the University of Illinois, Urbana-Champaign. INTSOY*, 10.
- Wittwer, Sylvan, et al. 1987. *Feeding a billion: Frontiers of Chinese agriculture*. East Lansing, Mich.
- Wolf, W. J., and J. C. Cowan. 1971. *Soybeans as a food source*. Cleveland, Ohio.
- Wright, H. E., Jr. 1968. Natural environment and early food production north of Mesopotamia. *Science* 161: 334-9.
- Yong, F. M., and B. J. B. Wood. 1974. Microbiology and biochemistry of the soy sauce fermentation. *Advanced Application Microbiology* 17: 157.

II.E.6 Sunflower

One of the most important of today's oil crops, the sunflower is a unique contribution of temperate North America to the world's major food plants. In addition to its superior oil, the seed of the sunflower is much appreciated as a food. Other parts of the plant were used for a variety of purposes by Native Americans. Today the species is also widely grown as an ornamental for its large showy heads.

Biology

Scientifically, the sunflower crop plant, known as *Helianthus annuus* var. *macrocarpus*, is a member of the family Asteraceae. It is an annual, is unbranched, grows from 1 to 3 meters tall, and bears a single large head up to 76 centimeters (cm) in diameter. Each head contains showy yellow sterile ray flowers and up to 8,000 smaller disk flowers. The latter produce the fruits, technically known as achenes, but commonly called seeds. The fruits, from 6 to 16 millimeters (mm) in length, contain a single seed.

In addition to the cultivated variety, the sunflower also includes branched, smaller-headed varieties (*Helianthus annuus* var. *annuus* and *Helianthus annuus* var. *lenticularis*) that are common as weeds or wild plants in North America from southern Canada to northern Mexico. Forms of the sunflower, particularly those with double flowers or red ray

flowers, are cultivated as ornamentals, but more so in Europe than in North America (Heiser 1976).

The genus *Helianthus*, native to North America, comprises 49 species and is divided into four sections (Schilling and Heiser 1981). The sunflower is placed in the section *Helianthus* along with 11 other annual species, all of which are diploid with 17 pairs of chromosomes. The silverleaf sunflower *Helianthus argophyllus* of Texas is the closest relative of the sunflower. The other sections include mostly perennial species, which may be diploid, tetraploid, or hexaploid. One of the perennials, the Jerusalem artichoke, is also cultivated as a food plant for its edible tubers.

Several of the species, both annual and perennial, are occasionally cultivated as ornamentals for their showy flowers (Heiser et al. 1969). *Helianthus* × *laetiflorus* and *Helianthus* × *multiflorus* are particularly widely cultivated for this purpose.

Origin and Early History

In prehistoric times, Native Americans in western North America collected the seeds of wild sunflowers for food, a practice continued by some until early in the twentieth century. In addition to eating the seeds, they also used the plants for pigments. The flowers were employed in ceremonies, the dried stems utilized in construction and for fuel, and various other parts exploited for medicinal purposes.

It has been postulated that the wild sunflower, originally confined to western North America, was a camp-following weed carried to central and eastern North America by the Native Americans, where it became an encouraged weed and, eventually, a deliberately cultivated plant. Mutations occurred, giving rise to monocephalic forms with larger seeds. The archaeological recovery of seeds in eastern North America indicates that the sunflower was domesticated before 1500 B.C. and that it was fairly extensively cultivated before 500 B.C. Some of these prehistoric sunflowers were in no way inferior to modern varieties in size. Early historical records reveal that the sunflower was also cultivated in Mexico and the American Southwest where it is still grown by the Hopi and Havasupai (Heiser 1951).

Along with sumpweed (*Iva annua*), a chenopod (*Chenopodium berlandieri*), and a squash (*Cucurbita pepo*), the sunflower was domesticated in eastern North America before the arrival of maize and other cultivated plants from Mexico (Smith 1989; Gayle 1990). By the time of early historical observations, it seems that maize had become the major crop of Native Americans, although the sunflower was still widely cultivated as a minor crop; sumpweed had disappeared entirely from cultivation.

Sometime after 1492, the sunflower was carried to Europe, the earliest introductions probably being in Spain. The first clear-cut reference to the sunflower is in the herbal of Rembert Dodoens in 1568, where it

is illustrated. The plant spread throughout much of Europe, reaching England before the end of the sixteenth century. The sunflower excited curiosity in some places because of its large size, but it was not an early success as a food plant. In fact, it was not until it reached Russia that the seeds became greatly appreciated.

The Holy Orthodox Church of Russia observed very strict dietary regulations during the 40 days of Lent and during Advent. In the early nineteenth century, the list of prohibited foods included nearly all of those rich in oil, but the sunflower, perhaps unknown to the church authorities, was not on the list. The Russian people eagerly adopted the sunflower seed, which could be used while still obeying the letter of the law (Gilmore 1931). The popularity of the sunflower was such that by midcentury or so Russia had become the world's leader in sunflower production.

Oil from sunflower seed was extracted in Russia from the early part of the nineteenth century onward. In addition to the heads, the stems were harvested as an important source of potash, some of which was exported (Putt 1978). Breeding to improve the crop plant commenced early in the twentieth century. Earlier-maturing and higher-yielding varieties were secured, and semidwarf varieties, which could be harvested mechanically, were developed. The oil content of the seeds was increased from 20 to more than 50 percent, and resistance to some pests and diseases was also incorporated into the plant.

Recent History

The two world wars led to great increases in sunflower production, particularly in Romania, Bulgaria, and Hungary. More recently, France and Spain have become major producers, but the former USSR has remained the world's leader and now produces more than half of the world's supply (Table II.E.6.1).

By far the most important sunflower-producing country in the Americas is Argentina, which became an exporter of the seeds early in the twentieth century. For many years, Argentina has been the world's second largest sunflower producer (see Table II.E.6.1).

Only recently, however, has the sunflower again become an important crop in its homeland. In both the United States and Canada, commercial sunflower cultivation was begun in the latter part of the nineteenth century, using varieties from Europe rather than native varieties. Indeed, North American production was, in part, the result of an introduction of improved varieties from the Soviet Union and the later adoption of higher-yielding hybrids (Putt 1978).

The sunflower was initially grown in the United States mostly for silage, but seed production has become increasingly important as people have replaced animal oils with those of plant origin in cooking. A quarter of a century ago, the sunflower, in some years, ranked second only to the soybean as one

Table II.E.6.1. *Sunflower production: Principal sunflower-producing countries in 1990 with 1960 figures for comparison (given in 1,000 metric tons)*

	1960	1990
Argentina	802	3,850
Bulgaria	279	365
Canada	17	115
China	-	1,500
France	3	2,314
Hungary	115	650
India	-	550
Italy	4	301
Morocco	-	160
Romania	529	556
South Africa	87	585
Spain	2	1,314
Turkey	128	900
United States	-	1,032
Former USSR	3,019	6,500
Former Yugoslavia	114	402
World	5,350	22,072

Source: FAO Production Yearbook, Vol. 14 (1961) and Vol. 44 (1991).

of the world's major oil crops. Lately, however, increased production of canola from rapeseed, particularly in Europe and Canada, has allowed this crop to take over the second position in oil production.

Agronomy

The sunflower is primarily a temperate zone crop. It has some tolerance to both high and low temperatures, and temperature may have an effect on oil composition. It is usually grown without irrigation, and although not highly drought tolerant, it often gives satisfactory yields when most other crops are damaged by drought. Sunflowers will grow in a variety of soil types, from clay to somewhat sandy, and respond well to fertilization, which may be done through soil or foliar treatments. Nitrogen is most often the limiting factor in their growth.

The plants may produce toxic residues in the soil so that crop rotation is essential for good growth. Weed control is also necessary, and herbicides are usually employed. In the northern United States, seeds are planted in late April or May. Later plantings affect oil composition adversely. The plants are ready for harvest in four months at which time considerable moisture is still present in the seeds. Thus, drying, usually by forcing air over the seeds, is required before the seeds are stored or transported (Robinson 1978; Schuler et al. 1978).

Pests and Diseases

Like most crops, the sunflower is subject to a number of insect pests, along with fungal and bacterial diseases. All parts of the plant may be attacked by

insects. The most serious pest in the United States is the sunflower moth, which infests the head (Schulz 1978). Downy mildew, a rust, and *Sclerotinia* white mold are among the main pathogens that damage the crop (Zimmer and Hoes 1978). Sunflower seeds are a favorite food of many birds, and at times up to 30 percent of the crop is lost to them. The chief culprit in the United States is the blackbird. One method of reducing the loss to birds is to harvest the crop early while the moisture content of the seeds is still high (Besser 1978).

Oil and Protein

The sunflower has excellent nutritional properties. The oil contains high levels of linoleic acid, moderate levels of linolenic acid, and less than 15 percent fatty acids (Dorrell 1978). Extraction of the oil is by solvents after the kernel is crushed or rolled and then cooked. The oil is used for margarine and shortening, as well as for cooking and salad oils. At present, it is little used for industrial purposes but has potential for a number of such applications.

The seed meal left after the extraction of oil serves as animal feed. It has about 28 percent protein and a good balance of amino acids, although it is low in lysine. The hulls, another by-product of the extraction process, are ground and used as filler or roughage in livestock feed in the United States. In other countries, the hulls are also used as fuel (Doty 1978).

Other Uses of the Seed

Although most of the world's sunflower crop goes into oil production, some is used for confectionery food and for feed for birds and pets (Lofgren 1978). The varieties employed for these purposes generally have larger seeds with lower oil content than those grown for oil extraction. The confectionery seeds are used as snack foods, in salads, and in bakery goods. The seeds are frequently shelled, roasted, and salted before reaching the consumer. Large amounts are sold in the shell in the United States for the feeding of wild birds.

Breeding

In the middle of the twentieth century it was realized that utilization of hybrid vigor could produce great increases in the yield of sunflowers. Because of the nature of the flower, the only practical way of making hybrids would be through the employment of cytoplasmic male sterility, and in 1968, Patrice Leclercq of France announced the discovery of cytoplasmic male sterility in crosses of sunflower with *Helianthus petiolaris* (Fick 1978). Fertility restorers were subsequently identified by a number of investigators, and the commercial production of hybrids was soon under way. Yield increases of up to 20 percent resulted, making the sunflower more competitive with a number of

other oil crops. Leclercq's cytoplasmic male sterility continues to be used for the creation of hybrids, but other sources of it are now available.

Sunflower breeding through traditional methods is being pursued in France, the United States, Russia, and a few other countries. In addition to increased yields, the principal objectives are to secure resistance to diseases and pests and to improve the oil content and composition. Both intra- and interspecific hybridization are being employed. The wild annual species have proved to be a particularly valuable source of disease resistance. The perennial species have, thus far, been little utilized.

Germ Plasm Reserves

Large seed collections that serve as a valuable reserve of genetic variation are maintained in Russia, the United States, and several other countries. Collections of the wild species are also available, and more than 2,000 accessions have been assembled in the United States (Seiler 1992). Wild *Helianthus annuus* still occurs naturally in large areas of North America, and several other annual species, particularly *Helianthus petiolaris*, have rather extensive distributions. Some of the other annual species occupy more restricted areas, but only one of them, *Helianthus paradoxus*, appears in imminent danger of extinction. Most of the perennial species are still readily available from wild populations in North America.

Charles B. Heiser, Jr.

Bibliography

- Besser, Jerome F. 1978. Birds and sunflowers. In *Sunflower science and technology*, ed. Jack F. Carter, 263-78. Madison, Wis.
- Dorrell, D. Gordon. 1978. Processing and utilization of oilseed sunflower. In *Sunflower science and technology*, ed. Jack F. Carter, 407-40. Madison, Wis.
- Doty, Harry O., Jr. 1978. Future of sunflower as an economic crop in North America and the world. In *Sunflower science and technology*, ed. Jack F. Carter, 457-88. Madison, Wis.
- Fick, Gerhardt N. 1978. Breeding and genetics. In *Sunflower science and technology*, ed. Jack F. Carter, 279-338. Madison, Wis.
- Gayle, Fritz J. 1990. Multiple pathways to farming in pre-contact eastern North America. *Journal of World Prehistory* 4: 387-435.
- Gilmore, Melvin R. 1931. Plant vagrants in America. *Papers of the Michigan Academy of Science, Arts and Letters* 15: 65-78.
- Heiser, Charles B. 1951. The sunflower among the North American Indians. *Proceedings of the American Philosophical Society* 98: 432-448.
1976. *The sunflower*. Norman, Okla.
- Heiser, Charles B., Dale M. Smith, Sarah B. Clevenger, and William C. Martin, Jr. 1969. The North American sunflowers (*Helianthus*). *Memoirs of the Torrey Botanical Club* 22: 1-218.
- Lofgren, James R. 1978. Sunflower for confectionery food, birdfood, and pet food. In *Sunflower science and technology*, ed. Jack F. Carter, 441-56. Madison, Wis.
- Putt, Eric D. 1978. History and present world status. In *Sunflower science and technology*, ed. Jack F. Carter, 1-29. Madison, Wis.
- Robinson, Robert G. 1978. Production and culture. In *Sunflower science and technology*, ed. Jack F. Carter, 89-113. Madison, Wis.
- Schilling, Edward E., and Charles B. Heiser. 1981. Infrageneric classification of *Helianthus*. *Taxon* 30: 293-403.
- Schuler, Robert T., H. J. Haring, V. L. Hofman, and D. R. Lundstrom. 1978. Harvesting, handling, and storage of seed. In *Sunflower science and technology*, ed. Jack F. Carter, 145-67. Madison, Wis.
- Schulz, J. T. 1978. Insect pests. In *Sunflower science and technology*, ed. Jack F. Carter, 169-223. Madison, Wis.
- Seiler, Gerald. 1992. Utilization of wild sunflower species as sources of genetic diversity for improvement of cultivated sunflower. *Field Crops Research* 30: 195-230.
- Smith, Bruce. 1989. Origins of agriculture in eastern North America. *Science* 246: 1566-71.
- Zimmer, David E., and J. A. Hoes. 1978. Diseases. In *Sunflower science and technology*, ed. Jack F. Carter, 225-62. Madison, Wis.



II.F

Trading in Tastes

II.F.1 🐉 Spices and Flavorings

Plants possess a wealth of different chemical ingredients, ranging from substances with simple structures to very complicated ones, such as terpene or benzoic derivatives. Some are poisonous, others are important raw materials in biochemistry and medicines, while still others are responsible for the appetizing odors we identify with certain foods.

Although the majority of spices are derived from plants, the most important spice is a mineral. Salt has been mined for culinary use (and perhaps, more importantly, as a food preservative) for more than 2,500 years, as well as secured at the seaside by the evaporation of seawater. Saltwater creatures, such as small herrings or other salty-tasting fish, have also been used as spices.

Among the spices of plant origin are such widely used spice plants as pepper and ginger. Other plants are frequently regarded as a flavoring, such as onions, peppers, carrots, and celery. Moreover, there are herbs, such as dandelions and daisies, which in the past were used as spices, as can be seen in old recipes.



Coriander

Although not normally considered to be spices, certain fruits and nuts, such as rowanberry, cranberry, and hazelnut, are also occasionally recommended in recipes as flavorings. In addition, plant stems, flowers, seeds, fruits, leaves, roots, and even pollen grains have been employed as spices.

Although all spices do not necessarily figure significantly in nutrition, there are other sound reasons for cooking with them, not the least of which are the aromatic ingredients they contain that can influence the taste of food. But spices also have chemical ingredients that aid digestion and play a role in food conservation (as does salt) by inhibiting bacterial growth and rendering the food unpalatable for microbes and insects. Until relatively recently, most spices had been luxury items, and their usage was an indicator of well-established culinary traditions within a relatively long-standing (and presumably highly developed) culture (Stobart 1977; Küster 1987).

The Landscapes of Spices

Spice plants occur in all parts of the world, but some regions, such as the tropics and subtropics, have an abundance of them, reflecting the greater diversity of plant species there. In the subtropics, the number of spice plants was increased by an extensive grazing of grasslands over millennia. Thus, for example, on the steppe landscapes of the Near East, North Africa, and the Mediterranean, cattle, goats, and sheep preferred to graze on mild, nonpoisonous plants without spines and thorns. Such selective pressure on the plant life of a region eventually produced the kind of subtropic grazing areas observed today, one that is characterized by a wealth of spice plants, albeit often spiny ones that are sometimes poisonous if taken in overly large quantities. Thus, the aromas of thyme, lavender, and sage are always present in the Mediterranean air and in the atmosphere of similar landscapes in Asia, as well as in North American areas such as California, Texas, and Florida, where such plants have been introduced by human migrants.

Origins of Spices

It is often difficult to reconstruct the origin of certain spices because they have been introduced by humans to many habitations worldwide throughout the ages. In fact, it is sometimes difficult to determine whether a plant is indigenous or introduced to a specific area. The latter may well have occurred before written sources existed, but unfortunately, even in historic times individuals rarely wrote about the introduction of plants.

Some ancient scientists and writers, such as Aristotle, Virgil, and Pliny, mention plants in their manuscripts. But whether they personally knew of those they wrote about or whether they mentioned them only because other scholars in earlier texts had done so is another matter. An added complication is the difficulty of botanically equating the plants referred to by spice names in Greek or Latin texts with their modern equivalents. Put plainly, that a plant was present at a specific time in a specific region often can be determined only by an archaeobotanical analysis of plant materials derived from layers of accurately dated archaeological excavations.

Even when a plant is found among archaeological remains, however, it is often unclear that its use had been as a spice. This question can be definitively

answered only with written sources. On the other hand, if pepper grains are found far outside their South Asian area of origin – in a Roman settlement in Europe, for instance – then this strongly suggests the likelihood that there had been a trade in pepper grains. And such a trade of pepper grains makes sense only if one assumes that they were used as spices in ancient Rome.

Having stressed the difficulties of reconstructing the ancient history of spices, we should also emphasize that most of the as-yet unresolved questions surrounding that history are now being addressed in the interdisciplinary work of botanists, archaeologists, and philologists. Such work has already delineated the most likely areas of the origin of the various spices, as listed in Table II.F.1.1 – a table that also shows that all of the well-known ancient civilizations used a complement of typical spices. Thus, ginger, star anise, and soy were known in early Chinese culture, whereas black pepper was commonly used by ancients in India. In the basins of the Euphrates and Tigris rivers, licorice and coriander have been cultivated since early times; ancient cultures of the Mediterranean used sage and anise. And, of course, vanilla, chocolate, and chilli peppers from the New World have become well-known gifts of the Mesoamerican civilizations.

Table II.F.1.1. *The origin of spices used historically and today*

Spice	Latin name	Area of origin	Time of oldest cultivation/consumption	Parts of plants taken as spice
Ajowan	<i>Trachyspermum ammi</i>	NE Africa, S Asia	Unknown	Fruits
Alexanders	<i>Smyrniolum olusatrum</i>	Mediterranean	Antiquity?	Leaves
Allspice	<i>Pimenta dioica</i>	Central America	Pre-Colombian	Fruits
Almond	<i>Amygdalus communis</i>	(W) Asia	2d millennium B.C.	Seeds
Angelica	<i>Angelica archangelica</i>	N Europe	Medieval	Stems, roots
Anise	<i>Pimpinella anisum</i>	S Europe	Bronze Age?	Fruits
Anise-pepper	<i>Xanthoxylum piperitum</i>	E Asia	Unknown	Fruits
Asafoetida	<i>Ferula asafoetida</i>	Persia, Afghanistan	2,000 B.C.	Dried mill
Avens	<i>Geum urbanum</i>	Widespread	Unknown	Roots
Balm	<i>Melissa officinalis</i>	E Mediterranean	Antiquity	Leaves
Basil	<i>Ocimum basilicum</i>	S Asia (India?)	Antiquity	Leaves
Bay	<i>Laurus nobilis</i>	Mediterranean	Antiquity	Leaves
Bergamot	<i>Monarda didyma</i>	N America	Unknown	Leaves
Black caraway	<i>Pimpinella saxifraga</i>	Europe, W Asia	Unknown	Leaves
Bog myrtle	<i>Myrica gale</i>	Northern Hemisphere	Medieval	Leaves
Borage	<i>Borago officinalis</i>	(W) Mediterranean	Middle Ages	Leaves, flowers
Burnet	<i>Sanguisorba</i> sp.	Europe	Unknown	Leaves
Calamus	<i>Acorus calamus</i>	SE Asia, N America	2d millennium B.C.	Roots
Caper	<i>Capparis spinosa</i>	Mediterranean, S,W Asia	2d millennium B.C.	Buds
Caraway	<i>Carum carvi</i>	Europe	Unknown	Fruits
Cardamom	<i>Elettaria cardamomum</i>	Tropical S Asia	3,000 B.C.	Seeds
Cassia	<i>Cinnamomum aromaticum</i>	China	3d millennium B.C.	Parts of bark
Cayenne pepper	<i>Capsicum frutescens</i>	Tropical America	Unknown	Fruits
Celery	<i>Apium graveolens</i>	Widespread at seashores	Neolithic	Fruits, leaves
Chervil	<i>Anthriscus cerefolium</i>	SE Europe, W Asia	Antiquity	Leaves
Chive	<i>Allium schoenoprasum</i>	N hemisphere	Unknown	Leaves
Chocolate	<i>Theobroma cacao</i>	Tropical America	Unknown	Seeds
Cinnamon	<i>Cinnamomum verum</i>	Ceylon	Antiquity	Parts of bark
Citron	<i>Citrus medica</i>	SE Asia	Unknown	Fruit of skin

Table II.F.1.1. (Continued)

Spice	Latin name	Area of origin	Time of oldest cultivation/consumption	Parts of plants taken as spice
Clary	<i>Salvia sclarea</i>	Mediterranean, W Asia	Antiquity?	Leaves
Clove	<i>Syzygium aromaticum</i>	Molukian Islands	Antiquity	Buds
Coriander	<i>Coriandrum sativum</i>	E Mediterranean, W Asia	Neolithic?	Fruits
Costmary	<i>Chrysanthemum balsamita</i>	W Asia	Unknown	Leaves
Cowslip	<i>Primula veris</i>	Eurasia	Unknown	Leaves, flowers
Cress	<i>Lepidium sativum</i>	NE Africa, W Asia	Unknown	Leaves
Cumin	<i>Cuminum cyminum</i>	E Mediterranean, W Asia	C. 3,000 B.C.	Fruits
Curry leaf	<i>Murraya koenigii</i>	S Himalaya	Unknown	Leaves
Dandelion	<i>Taraxacum officinale</i>	Europe	Unknown	Leaves
Dill	<i>Anethum graveolens</i>	S Europe? Asia?	3,000 B.C.	Leaves
Elecampane	<i>Inula helenum</i>	SE Europe	Antiquity	Roots
Fennel	<i>Foeniculum vulgare</i>	Mediterranean	3,000 B.C.	Fruits
Fenugreek	<i>Trigonella foenum-graecum</i>	E Mediterranean	3,000 B.C.	Seeds
Galangal	<i>Alpinia galanga/officinarum</i>	SE Asia	Unknown	Roots
Garlic	<i>Allium sativum</i>	Central Asia	2d millennium B.C.	Bulbs
Garlic mustard	<i>Alliaria petiolata</i>	Europe, W Asia, N Africa	Unknown	Leaves
Gentian	<i>Gentiana lutea</i>	Europe	Antiquity?	Roots
Geranium	<i>Pelargonium</i> sp.	S Africa	Unknown	Oil of leaves
Ginger	<i>Zingiber officinale</i>	SE Asia	2d millennium B.C.	Roots
Grains of paradise	<i>Aframomum melegueta</i>	W Africa	Unknown	Seeds
Ground ivy	<i>Glechoma hederacea</i>	Eurasia	Unknown	Leaves
Hazel nut	<i>Corylus avellana</i>	Eurasia	Mesolithic	Seeds
Hop	<i>Humulus lupulus</i>	Europe	Medieval	Fruits
Horehound	<i>Marrubium vulgare</i>	Asia, S Europe	Antiquity	Leaves
Horseradish	<i>Armoracia rusticana</i>	SE Europe, W Asia	Unknown	Roots
Hyssop	<i>Hyssopus officinalis</i>	Eurasian mountains	Unknown	Leaves
Indian cress	<i>Tropaeolum majus</i>	Tropical S America	Unknown	Leaves
Juniper	<i>Juniperus communis</i>	N Hemisphere	Unknown	Fruits
Lady's smock	<i>Cardamine pratensis</i>	N Hemisphere	Unknown	Leaves
Lavendar	<i>Lavandula officinalis</i>	S Europe	Unknown	Leaves
Lemon	<i>Citrus limon</i>	SE Asia	Unknown	Fruit skin
Lemon grass	<i>Cymbopogon citratus</i>	SE Asia, Ceylon	Unknown	Leaves
Lime	<i>Citrus aurantium</i>	Tropical SE Asia	Unknown	Fruit skin
Liquorice	<i>Glycyrrhiza glabra</i>	E Mediterranean, Asia	2,000 B.C.	Roots
Lovage	<i>Levisticum officinale</i>	Persia?	Antiquity	Leaves
Mace, nutmeg	<i>Myristica fragrans</i>	Mollukian Islands	2d millennium B.C.	Seeds, seed cover
Marigold	<i>Calendula officinalis</i>	Mediterranean?	Unknown	Flowers
Marjoram	<i>Origanum majorana</i>	E Mediterranean	Antiquity?	Leaves
Mastic	<i>Pistacia lentiscus</i>	Mediterranean, Canary Islands	Antiquity	Resin
Melilot	<i>Melilotus officinalis</i>	N Hemisphere	Unknown	Leaves
Milfoil	<i>Achillea millefolium</i>	Europe, N Asia	Unknown	Leaves
Mint	<i>Mentha</i> sp.	Widespread	Unknown	Leaves
Mugwort	<i>Artemisia vulgaris</i>	N Hemisphere	Unknown	Leaves
Mustard	<i>Brassica</i> sp., <i>Sinapis alba</i>	Eurasia	Antiquity	Seeds
Myrtle	<i>Myrtus communis</i>	Mediterranean	Antiquity	Leaves, fruits
Nasturtium	<i>Tropaeolum majus</i>	S America	Unknown	Leaves
Nigella	<i>Nigella sativa</i>	SE Europe, W Asia	Antiquity	Seeds
Olive	<i>Olea europaea</i>	(E) Mediterranean?	Neolithic?	Fruits
Onion	<i>Allium cepa</i>	Central Asia	3d millennium B.C.	Bulbs
Orange	<i>Citrus sinensis</i>	Tropical SE Asia	Unknown	Fruit skin
Oregano	<i>Origanum vulgare</i>	Eurasia	Antiquity	Leaves
Parsley	<i>Petroselinum crispum</i>	Mediterranean	3,000 B.C.	Leaves
Parsnip	<i>Pastinaca sativa</i>	Europe, W Asia	Antiquity?	Fruits, seeds
Peanut	<i>Arachis hypogaea</i>	Brazil	Unknown	Seeds
Pepper	<i>Piper nigrum</i>	Tropical SE Asia	Prehistoric India	Seeds
Peppers	<i>Capsicum annum</i>	Tropical America	Unknown	Fruits, seeds
Pine nut	<i>Pinus pinea</i>	Mediterranean	Before 2,000 B.C.	Seeds
Pistachio nut	<i>Pistacia vera</i>	W Asia	Neolithic	Seeds
Pomegranate	<i>Punica granatum</i>	E Mediterranean, W Asia	3d millennium B.C.	Fruits
Poppy	<i>Papaver somniferum</i>	(W) Mediterranean	Neolithic	Seeds

(continued)

Table II.F.1.1. (Continued)

Spice	Latin name	Area of origin	Time of oldest cultivation/consumption	Parts of plants taken as spice
Purslane	<i>Portulaca oleracea</i>	S Asia?	2,000 B.C.	Leaves
Plantain	<i>Plantago lanceolata</i>	Eurasia	Unknown	Leaves
Rocket	<i>Eruca sativa</i>	Mediterranean, W Asia	Antiquity?	Seeds
Rose	<i>Rosa</i> sp.	Widespread	Unknown	Petals, fruits
Roselle	<i>Hibiscus sabdariffa</i>	Tropical S Asia	Unknown	Calyx
Rosemary	<i>Rosmarinus officinalis</i>	Mediterranean	Antiquity	Leaves
Rowan	<i>Sorbus aucuparia</i>	Europe	Unknown	Fruits
Rue	<i>Ruta graveolens</i>	Balkans, Italy	Antiquity	Leaves
Safflower	<i>Carthamus tinctorius</i>	W Asia	Unknown	Flowers
Saffron	<i>Crocus sativus</i>	Greece? Near East?	Unknown	Pollen
Sage	<i>Salvia officinalis</i>	Mediterranean	Antiquity	Leaves
Samphire	<i>Critblum maritimum</i>	European coasts, Canary Islands	Antiquity	Leaves
Savory	<i>Satureja hortensis</i>	Mediterranean	Roman	Leaves
Screwpine	<i>Pandanus tectorius</i>	India, Ceylon	Unknown	Mainly flowers
Sea buckthorn	<i>Hippophaë rhamnoides</i>	N Eurasia	Unknown	Fruits
Sesame	<i>Sesamum indicum</i>	E Africa? India?	Ancient Orient	Seeds
Silphion	? (extinct)	N Africa?	Antiquity	?
Sorrel	<i>Rumex</i> sp.	Widespread	Unknown	Leaves
Southernwood	<i>Artemisia abrotanum</i>	SE Europe, W Asia	Antiquity	Leaves
Soy	<i>Glycine max</i>	SE Asia	Unknown	Seeds
Star anise	<i>Illicium verum</i>	SE Asia	2,000 B.C.	Fruits
Stonecrop	<i>Sedum reflexum</i>	Europe	Unknown	Leaves
Sugar beet	<i>Beta vulgaris</i>	Europe	18th century	Roots
Sugercane	<i>Saccharum officinarum</i>	SE Asia	Antiquity	Stalks
Sumac	<i>Rhus coriaria</i>	Mediterranean, W Asia	Antiquity	Fruits
Sunflower	<i>Helianthus annuus</i>	Central America	Unknown	Seeds
Sweet cicely	<i>Myrrhis odorata</i>	Europe	Unknown	Fruits
Sweet woodruff	<i>Asperula odorata</i>	Europe	Unknown	Leaves
Tamarind	<i>Tamarindus indica</i>	Tropical Africa, (India?)	2,000 B.C.	Parts of fruits
Tansy	<i>Tanacetum vulgare</i>	Eurasia	Unknown	Leaves
Tarragon	<i>Artemisia dracunculus</i>	Asia, N America	2d millennium B.C.	Leaves
Thyme	<i>Thymus vulgaris</i>	Mediterranean	Unknown	Leaves
Turmeric	<i>Curcuma longa</i>	Tropical SE Asia	Unknown	Roots
Vanilla	<i>Vanilla planifolia</i>	Tropical America	Pre-Columbian	Fruits
Violet	<i>Viola odorata</i>	Mediterranean, S England	Antiquity?	Flowers
Walnut	<i>Juglans regia</i>	E Mediterranean?	Antiquity	Seeds
Watercress	<i>Nasturtium officinale</i>	Widespread	Unknown	Leaves
White cinnamon	<i>Canella alba</i>	Central America	Unknown	Bark
Wild thyme	<i>Thymus serpyllum</i>	Europe, W Asia	Unknown	Leaves
Wood sorrel	<i>Oxalis acetosella</i>	N Hemisphere	Unknown	Leaves
Wormwood	<i>Artemisia absinthium</i>	Eurasia	Antiquity	Leaves

Spices as Elements of an Urban Lifestyle

It would seem that those emerging from the Old World Neolithic Revolution who pioneered in employing most of the spices were not rural peasants but rather urban dwellers: Historical evidence suggests that it was only in those cultures which gave rise to concentrated settlements that spice use seems to have been known.

In most cases, that use was aimed not so much at making the food more tasty but rather at masking the taste of food that was no longer fresh, or even

spoiled. Spoilage was not a major problem with cereal grains and legumes. But it was more of a problem with the less easy-to-convert oil plant seeds and, of course, a major one with fish, fowl, meat, and animal products such as milk.

Although early humans appear to have been scavengers, it was presumably the ideal among their prehistoric hunter-gatherer successors to consume a hunted animal immediately after it had been killed. This, in turn, would have made the kill of any large animal a social event, with friends and relations joining the hunters in eating the meat while it was

still fresh and unspoiled. After everybody had been fed, however, there was probably little interest in the remaining scraps, and because meat conservation technology had not been developed, it was necessary to kill again before meat once more entered the diet.

A similar approach to meat consumption was, doubtless, the case in the first farming communities. But as these early farmers increasingly derived more and more of their nutrients from crops, meat became less important in their diets – much less than it had been among their hunter-gatherer forebears.

Yet the relative dietary importance of various foods shifted again when the first towns or townlike settlements arose. Town dwellers depended for their food on rural settlements, and this demand was more readily met by meat than by grain. Clearly it was far easier for rural producers to drive cattle and sheep, moving under their own power, to a slaughterhouse inside a town than it was to haul bulky, not to mention heavy, containers of grains and legumes to town storage areas from the countryside.

Meat, as a consequence, became an important part of the diet in urban settlements, which led, in turn, to technology that permitted the utilization of the entire carcass. On the one hand, each part of the carcass had food value, and on the other, such utilization avoided the problem of carcass disposal inside or near the settlements, with the attendant difficulties of odors, pests, and even disease associated with the rotting remains.

The first techniques of meat conservation were those of smoking plus salting to preserve meats from insects and microbes that cannot live in a salty environment. However, even salted meat was often spoiled, or at least tasted spoiled, by the time it was consumed, and it was discovered that the spoiled taste could be covered by the addition of pungent spices. Moreover, although people might fall ill from the consumption of spoiled meat, it was learned that other spices could help prevent these consequences by making digestion easier.

Because of such growing knowledge, the people of the towns that demanded meat also demanded spices, and this was especially true of the upper classes. It came to be regarded as sophisticated to mask the taste of spoiled meat and fat in cookery and, thus, spices became a valuable item of exchange.

Elsewhere, and considerably later, other populations that had come to depend heavily on fish learned to preserve them by creating spicy sauces to accompany them. And later in the Americas, spices, especially chilli peppers, enlivened fare that in many cases was essentially vegetarian in nature. Clearly, then, contact with the spice trade that developed in the Eurasian world was not a prerequisite for spice usage. On the other hand, those with money who enjoyed such contact could choose from a variety of seasonings.

The Beginning of the Spice Trade

The spice trade began as early as the trade in precious metals and jewels and dyes and silks – which, along with spices, were ancient items traded over the silk and spice roads in the Near East. Spices first moved between Mesopotamia and Egypt; later, Indian spices were traded to the Levant, and Mesopotamian spices reached the Indus valley (see Table II.F.1.1 for areas of origin). All important trade centers were connected by spice roads, with such trade generally practiced by nomadic peoples, who in later times were the Arabs.

As the spice roads of the Near Eastern deserts were extended to Levantine harbors, spices began traveling by ship to Athens, Rome, Massilia, and other parts of the ancient world. The Greeks and Romans, in turn, added a number of Mediterranean spices to the trade and began to vie for dominance of the trade. In fact, control of the spice trade was one factor in the warfare among the Greeks and Romans and Near Eastern peoples.

Roman consumption of spices from all over the known world was enormous. In fact, the Roman demand for the spice silphion, which was grown in Kyrene in Northern Africa, was so great that it was overexploited until it became extinct. Today silphion is known only through historical documents, and its conventional designation as a member of the Apiaceae plant family can only be a tentative one (Küster 1987). Moreover, because the Romans took spices to all places on the globe where Roman soldiers were garrisoned, Roman cookery became the first cuisine that (at the risk of argument from Chinese historians) might be called “international.” In addition, it was the Romans who established the first trade network across the known world to send spices, such as pepper, from Southeast Asia all the way to central Europe (Kucan 1984).

Use of Spices in Medieval Europe

During the European Middle Ages, Roman cookery and its use of spices had come to be well regarded, even in countries far more distant from the pepper-growing areas than ancient Rome had been. Spices, brought to western and northern Europe, were shipped along the western coast of Europe to England, Ireland, France, and the Netherlands. The Rhine river became an important shipping route as well. The Hanseatic merchants traded spices in the North Sea and the Baltic areas, and later, northern Italian and southern German merchants began to trade them overland and via the Alps. Harbor towns like Venice, Genoa, Lisbon, London, Dublin, Amsterdam, and Lübeck became wealthy as a result of the spice trade, as did inland towns like Constance, Augsburg, Nuremberg, and Cracow.

Indeed, at times, medieval European kings and emperors could not reign without the support of the spice merchants, who were very wealthy and exerted considerable political influence. The nickname "Pfeffersack" (pepper sack) was commonly used to refer to a merchant making very high profits from the spice trade. The use of black pepper, which could mask the scent of spoiled meat and fat, came to be regarded as a status symbol. Duke Karl of Burgogne, regarded in his day as the richest man in Europe, ordered 380 pounds of pepper for his wedding dinner in 1468 (Küster 1987).

Ordinary people, however, could not afford to buy spices from Southeast Asia, which thus remained in short supply despite the trading networks. For this reason, there were often attempts to adulterate spices, which became a serious kind of crime that could result in a sentence of death by burning or decapitation. All of this, in turn, provided a strong motivation for people to discover pungent plants that were available locally and might be substituted for expensive imported spices. Among such finds were juniper, gentian, and caraway – plants that were unpalatable to livestock and, consequently, commonplace in the intensively grazed lands surrounding medieval towns. Gradually their seeds, leaves, fruits, and roots found their way into local cookery.

In Search of New Trading Routes

The Crusades marked the beginning of a crisis in the worldwide spice trade, especially in central Europe, as the significant trade between Arabs and Venetian merchants was disrupted. This development, in turn, produced what might be loosely termed "scientific" research into the origin of spices in central and western Europe. At that time, knowledge of the subject consisted of half-true tales passed down by Greek and Roman writers. But now physicians and others concerned with human health began to think about the active ingredients in herbs and spices, and an argument can be made that a shortage of spices was one of the stimuli that led to the development of Renaissance science.

The spice shortage also contributed to improvements in navigation technology, as several European nations competed to discover sea routes to India. In other words, this early exploration was not so much to bring Western civilization and Christianity to other peoples as to profit from the trade in precious stones, metals, and spices. Similarly, at a later time, ancient civilizations in Africa and Asia were set upon by European colonialists, and one of the precipitating factors for such imperialistic adventures was the competition of spice merchants from western and southern Europe.

The Portuguese found a sea route from Europe to India via the circumnavigation of Africa, which allowed them to bypass Arab intermediaries and bring

such spices as pepper and cinnamon directly to Europe. They also discovered new spices, such as the "grains of paradise" (*Aframomum melegueta*) from western Africa, which they sold instead of pepper by bestowing upon the grains the enticing but misleading name "melegueta pepper."

Meanwhile, Columbus and others had hypothesized that another possible trading route to India lay to the west, and stimulated in part by the desire to find new sea routes for the spice trade, the Spaniards "discovered" "Western India," or America. However, the Europeans found not pepper but (chilli) peppers, white cinnamon instead of cinnamon, Indian cress instead of cress, as well as chocolate and vanilla. Such American spices became an important part of the wealth of the New World to be exploited by the Old.

By the start of the sixteenth century, western European countries were beginning to assert their control over world trade and politics, largely due to their quest for spices. Thus, the spice trade figured prominently in European colonialism and imperialism and, consequently, in shaping the world political order of today.

The Internationalization of Spices

Since the sixteenth century, and especially since the nineteenth century, many Europeans and Asians have migrated to the Americas, taking with them their culinary traditions, including their knowledge of spices and how to use them. Because such knowledge was subsequently exchanged among different cultural groups, the use of particular spices became increasingly less identified with the cuisines that had originally incorporated them.

First iceboxes and then refrigerators diminished the risk of food spoilage, and thus erased the importance of spices in masking the taste of food going bad – the paramount reason for seeking them in the first place. But if spices were less necessary in cookery for covering up unpleasant tastes, they did provide pleasant flavors of their own, which strongly established them in cookery traditions. And this, coupled with the mass production and marketing that has made them so inexpensive, has ensured their continued widespread usage. Indeed, it is possible today to buy virtually any spice in the supermarkets of the world at an affordable price.

Of course, some national or regional cuisines can still be identified by their traditional spices. Restaurants that specialize in various types of international cuisine have proliferated – especially since the end of colonialism. North Africans who migrated to France established Algerian and Moroccan restaurants in Paris. Many Indonesian restaurants can be found in the Netherlands. Chinese and Indian restaurants have become popular in Great Britain and in the United States – to provide merely a few examples.

Somewhat later, a different development took

place in Germany. During the 1950s and 1960s, a labor shortage in Germany induced many southern Europeans to venture northward seeking employment. And because of this migration, Italian, Balkan, and Greek restaurants with all their accompanying spices are now common in Germany.

Spices are so widely employed today that we scarcely notice their presence. Most understand and would agree that spices improve the taste of food and even aid digestion. But largely forgotten is the fact that the original purpose of spices was to mask bad tastes, rather than provide good ones; and almost completely forgotten is the role that the spice trade played in stimulating scientific thought during the Renaissance and the explorations and the empire building that followed.

Hansjörg Küster

Bibliography

- Foster, Nelson, and Linda S. Cordell. 1992. *Chilies to chocolate: Food the Americas gave the world*. Tucson, Ariz.
- Kucan, Dusanka. 1984. Der erste römerzeitliche Pfefferfund – nachgewiesen im Legionslager Oberaden (Stadt Bergkamen). *Ausgrabungen und Funde in Westfalen-Lippe* 4: 51–6.
- Küster, Hansjörg. 1987. Wo der Pfeffer wächst. *Ein Lexikon zur Kulturgeschichte der Gewürze*. Munich.
- Nabhan, Gary Paul. 1985. *Gathering the desert*. Tucson, Ariz.
- Stobart, Tom. 1977. *Herbs, spices and flavorings*. Harmondsworth, England.

II.F.2 🍷 Sugar

Sugar is the world's predominant sweetener. It satisfies the human appetite for sweetness and contributes calories to our diet. Sugar is used in cooking, in the preparation of commercially processed foods, and as an additive to drinks; it is also a preservative and fermenting agent. It sweetens without changing the flavor of food and drink. It is cheap to transport, easy to store, and relatively imperishable. These characteristics helped sugar to displace such sweeteners as fruit syrups, honey, and the sap of certain trees, the most famous of which is the North American maple.

Lack of data makes it difficult to establish when sugar became the principal sweetener in any given part of the world, but in every case this has occurred fairly recently. Illustrative are Europe and North America where it was only after 1700 that sugar was transformed from a luxury product into one of everyday use by even the poor. This took place as Brazil and the new West Indies colonies began producing sugar in such large quantities that price was significantly reduced. Lower prices led to increased consumption,

Sugar cane



which, in turn, fueled demand, with the result that the industry continued to expand in the Americas and later elsewhere in the tropical world.

Since the eighteenth century, the rise in the per capita consumption of sugar has been closely associated with industrialization, increased personal income, the use of processed foods, and the consumption of beverages to which people add sugar, such as tea, coffee, and cocoa. In addition, the relatively recent popularity of soft drinks has also expanded the use of sugar. Annual per capita sugar consumption is now highest in its places of production, such as Brazil, Fiji, and Australia, where it exceeds 50 kilograms (kg). Consumption in Cuba has been exceptionally high, exceeding 80 kg per capita around the beginning of the 1990s. Subsequently, consumption has fallen to a still very high 60 kg per person.

With an annual per capita consumption of between 30 and 40 kg, the countries that were first industrialized in western Europe and North America constitute a second tier of sugar consumers. The poorer countries of the world make up a third group where consumption is low. The figure for China is 6.5 kg, and it is even lower for many countries in tropical Africa. Such a pattern reflects both differences in wealth and the ready availability of sugar to those in the countries of the first group. In the Western industrialized world, concerns about the effects of sugar on health, as well as the use of alternatives to sugar – such as high-fructose corn syrup and high-intensity, low-calorie sweeteners – have stabilized and, in some

countries, lowered the use of sugar. Thus, it would seem that further expansion of the industry depends primarily on the poorer countries following the precedent of the richer ones by increasing consumption as standards of living improve. Secondarily, it depends on the ability of the sugar industry to meet competition from alternative sweeteners.

Sugar is the chemical sucrose that occurs naturally in plants. It is most richly concentrated in sugarcane and sugar beet, which are the sources of commercial sugar. Fully refined sugar, whether made from cane or beet, is pure sucrose, and the consumer cannot tell from which of the two plants it derives. But despite the identical end product, the sugarcane and the sugar beet industries differ greatly in methods of production and organization, and each has its own distinctive history and geography.

The Sugarcane Industry

The Raw Material

Sugarcane is a perennial grass of the humid tropics. It requires at least 1,000 millimeters of rain annually, which should fall year-round, although with irrigation the plant can also be grown in dry climates. Temperatures must be above 21° Celsius (C) for satisfactory growth, and the best results are obtained when the temperature exceeds 27° C. Cold temperatures, therefore, impose northern and southern limits on cultivation. Growth ceases when the temperature falls below 11–13° C; light frosts injure the plant, and a prolonged freeze will do serious damage. Sugarcane is tolerant of a wide range of soil conditions and grows well on both hillsides and flat land. With the mechanization of harvesting, a recent development that dates only from the 1950s, the industry has come to prefer flat land where the machines can function most effectively.

The principal parts of sugarcane are the roots and the stem that supports the leaves and the inflorescence. The stem in a mature plant can grow as high as 5 meters (m) and can be thick or thin, depending on the variety. Some stems have soft rinds and are easy to chew on; others have tough rinds, which makes for difficult milling. The color of the stems can range from green through shades of purple, and some varieties are leafy, whereas others are not. Despite these differences, the identification of varieties in the field is usually a matter for experts.

Commercial sugarcane is reproduced vegetatively. Until the late nineteenth century, the commercial varieties were thought to be infertile. Some are, although others set seed under certain climatic and day-length conditions. This discovery has been of basic importance for the breeding of new cane varieties, but commercial sugarcane is still reproduced in the traditional vegetative way. The stems have nodes spaced from 0.15 to 0.25 m apart, each of which contains root primordia and buds. A length of stem with at least one node is known variously as a sett, stem-cutting, or

seed-piece. When setts are planted, roots develop from the primordia and a stem grows from the bud. Stems tiller at the root so that the bud in each sett produces several stems.

The first crop, known as plant cane, matures in 12 to 18 months, depending on the climate and variety of cane. The roots, left in the ground after the harvest, produce further crops known as ratoons. Some varieties produce better ratoons than others. As a perennial plant with a deep root system and with good ground coverage provided by the dense mat of stems and leaves, sugarcane protects the soil from erosion. Given adequate fertilizer and water, it can flourish year after year, and there are parts of the world in which sugarcane has been a cash crop for centuries.

All wild and domesticated varieties of sugarcane belong to the genus *Saccharum*, one of the many subdivisions of the large botanical family of *Gramineae*. Because these varieties interbreed, the genus has become a very complex one. Authorities recognize species of *Saccharum*: *S. robustum* Brandes and Jeswiet ex. Grassl., *S. edule* Hassk., *S. officinarum* L., *S. barberi* Jeswiet, *S. sinense* Roxb. emend. Jeswiet, and *S. spontaneum*. Speciation, however, of the genus is still in dispute.

The status of *S. edule* as a distinct species is problematic, and an argument exists for conflating *S. barberi* and *S. sinense*. There is also discussion as to the places of origin of the species. *S. barberi*, *S. sinense*, and *S. spontaneum* occur widely in southern Asia and may have originated there. New Guinea is thought by many to have been the place of origin of the other three species, although possibly *S. officinarum* evolved close by in the Halmahera/Celebes region of present-day Indonesia. Four of the species were cultivated for their sugar: *S. barberi* and *S. sinense*, respectively, in India and China; and *S. edule* and *S. officinarum* in New Guinea. *S. robustum* is too low in sucrose to be a useful source of sugar, but its tough fiber makes it valued in New Guinea for fencing and roofing. *S. officinarum* may have been cultivated as an important food for pigs in prehistoric New Guinea (Stevenson 1965: 31–2; Barnes 1974: 40–2; Blackburn 1984: 90–102; Daniels and Daniels 1993: 1–7).

S. officinarum is the species of basic importance to the history of the sugarcane industry. In New Guinea its evolution into an exceptionally sweet cane led to wide diffusion throughout the Pacific islands and eastward through southern Asia to Mediterranean Europe and America. Until the 1920s, nearly all the cane sugar that entered international commerce came from one or another variety of this species.¹ The varieties of *S. officinarum* are known collectively as noble canes, a name the Dutch gave them in the nineteenth century in appreciation of their importance to the sugar industry in Java.

Disease brought the era of the noble canes to an end. In the mid-nineteenth century, Bourbon cane, the standard commercially cultivated variety in the

Americas, had suffered occasional outbreaks of disease, but in the West Indies in the 1880s, disease in Bourbon cane became general and caused a serious reduction in yields. To use the jargon of the sugar trade, this variety "failed."

In the 1880s, disease also devastated the cane fields of Java. The initial response of planters was to replace the diseased cane with other "noble" varieties, but because these too might "fail," this solution was seen at best to be a temporary one. The long-term answer lay in breeding new varieties that would be resistant to disease and rich in sucrose. Cane-breeding research began simultaneously in the 1880s at the East Java Research Station (Proefstation Oost Java) and at Dodds Botanical Station in Barbados. Initially, researchers worked only with *S. officinarum*. Success came slowly, but during the first years of the twentieth century, newly bred varieties began to replace the naturally occurring ones, and substitution was complete in most regions by the 1930s.

In a second phase of research, which benefited from developments in genetics, attractive features from other species of sugarcane were combined with varieties of *S. officinarum*. This process, known as "nobilization," resulted in even further improvements in disease resistance and sucrose content. Nobilized varieties entered cultivation during the 1920s and gradually replaced the first generation of bred varieties (Galloway 1996).

The sugarcane industry has now been dependent on cane breeding for a century, and research remains necessary to the success of the industry. Because varieties "fail," reserve varieties must be on hand to replace them. Some of the aims of breeding programs are long-standing: resistance to disease, insects, and animal pests; suitability to different edaphic and climatic conditions; better ratooning; and high sucrose content. In recent years, additional considerations have come to the fore. Canes have to be able to tolerate herbicides, and in some countries they have to meet the needs of mechanical harvesters.

Cultivation and Harvest

Success in cane breeding gave the sugarcane industry a much improved raw material that enables it to meet more easily a basic requirement of survival in a competitive world, which is the sustainable production of cane rich in sucrose in different climatic and soil conditions. Cane breeding has also added flexibility in dealing with the constant of how best to manage the cane fields with a view to the needs of the mill and factory. A long harvest season is preferable to a short one because it permits the economic use of labor and machinery. Fields of cane, therefore, must mature in succession over a period of months. This can be achieved by staggering the planting of the fields and by cultivating a combination of quick-maturing and slow-maturing varieties. Cane ratoons complicate the operation because they mature more quickly than the plant cane.

Over the years, however, ratoon crops gradually decline in sugar yield, and eventually the roots have to be dug up and the field replanted. Breeding can improve the ratooning qualities of cane. The number of ratoon crops taken before replanting is a matter of judgment, involving the costs of planting and the yield of the crops. Practice over the centuries has ranged from none at all to 15 and even more. Where numerous ratoon crops are the custom, the introduction of new varieties is necessarily a slow process. Planting is still largely done by hand, although planting by machine has increased.

The harvesting of cane presents another set of problems. The cane must be carefully cut close to the ground (the sucrose content is usually highest at the base of the stems) but without damage to the roots. The leaves and inflorescence should not go to the mill where they will absorb sugar from the cane and so reduce the yield. Manual workers can cut carefully and strip the stems, but the work is arduous and expensive. The mechanization of harvesting has been made easier by the breeding of varieties that achieve uniform height and stand erect. The machines can cut and top (remove the inflorescence) with little waste.

Because it is very important to send clean cane to the mill with a minimum of trash, soil, or other debris, preharvest burning of the cane has become a common practice. This removes the living leaves as well as the dead, which lie on the ground and are known as "trash." Burning prepares a field for harvesting and does not damage the stems or reduce the yield of sugar, provided the stems reach the mill within 24 hours. But this last requirement is not new: Breeding has not been able to alter the fact that stems of cane are perishable and, once cut, must be milled quickly to avoid loss of juice and, hence, of sucrose and revenues. The perishable nature of the stems and the expense of transporting them still demand that the mills be located close to the fields where the cane is grown.

Where the owners of the mills also own the fields that supply the cane, coordination of cultivation and harvesting is a relatively easy matter, compared to a situation in which independent cane growers do the supplying. There are mills that rely on hundreds of cane growers, each cultivating a few hectares. In addition to logistical issues, the price the growers are to receive for their cane is a constant source of dispute (Blackburn 1984: 136-288; Fauconnier 1993: 75-132).

From Mill to Refinery

The aim in the mill is to extract the maximum amount of sucrose from the stems; the aim in the factory (usually attached to the mill) is to make maximum use of the sucrose to produce sugar quickly and efficiently. The opaque, dark-green juice that contains the sucrose flows from the mill to the factory where it is first heated and clarified, then condensed by boiling to a thick syrup in which sugar crystals form. This

mix, known as a "massecuite," is then separated into sugar crystals and molasses. Early in the history of the industry, both milling and manufacture involved rather simple processes that by modern standards not only were slow and wasteful but also led to rather crude results. Now the milling and manufacture of sugar involve a series of highly sophisticated engineering and chemical operations and take place in what are, in fact, large industrial complexes. The daily grinding capacity of modern mills varies enormously, with the largest exceeding 20,000 metric tons (Chen 1985: 72).

Innovations in processing have never entirely replaced the old methods, and in India, Brazil, and many other countries modern factories coexist with traditional ones. The industry draws a distinction between the products of the two, based on the method of separating the crystals from the molasses. In traditional mills, the massecuite is placed in upright conical pots, and over a period of days, if not weeks, the molasses drains through a hole at the base of the cone, leaving a sugar loaf. In a modern mill, separation is accomplished rapidly in a centrifugal machine that was first employed in sugar factories in the mid-nineteenth century. The terms "centrifugal" and "noncentrifugal" are used in the sugar trade as a shorthand, referring to the products of very different processes of manufacture.

Noncentrifugal sugars are characteristically made for local markets by rural entrepreneurs with simple equipment and little capital. They appear for sale yellowed by molasses, and sometimes still moist, and compete in price with centrifugal sugar. They appeal to people who are especially fond of their taste. The noncentrifugal sugars include the *gur* of India, the *jaggery* of Nigeria, the *rapadura* of Brazil, and the *panela* of Colombia. (Indian *kbandsari*

sugar, an interesting anomaly, is produced by traditional methods, except that the crystals are separated in hand-held centrifuges.) The often remote location and modest scale of many noncentrifugal factories means that a proportion of world cane sugar production goes unreported.

Centrifugal sugar, the product of the modern factories, by contrast, enters world trade. The International Sugar Organization, as well as governments and sugar traders, track production; thus, there are detailed, country-by-country, annual statistics on production and consumption (Figure II.F.2.1). Commercial sugars are classified according to their purity, which is measured in degrees of polarization (pol.) with a polarimeter; purity ranges from 0° pol., signifying a total absence of sugar (as in distilled water, for example), to 100° pol., indicating pure sucrose. The basic centrifugal sugar of commerce is known as raw sugar, which, by international agreement, must have a polarization of at least 96°. Raw sugars are not quite pure sucrose and are intended for further processing, meaning that they are to be refined.

Refining is the final stage in the manufacture of sugar. The result is pure sucrose, with a pol. of 100°. The process of refining involves melting the raws, removal of the last impurities, and recrystallization of the sucrose under very careful sanitary conditions. The raws lose a very minor percentage of their weight in the transformation to refined sugar, the amount depending on the pol. of the raws and the technical ability of the refinery (Chen 1985: 634–6). This loss and the expense of refining are built into the selling price of refined sugar.

Refineries are located close to the market where the sugar is to be consumed, rather than the fields where the cane is grown. In the Western world, this

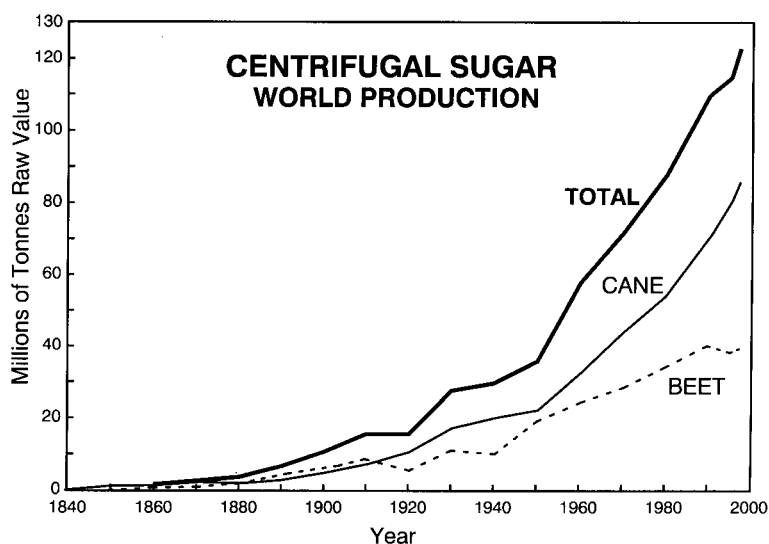


Figure II.F.2.1. Centrifugal sugar; world production. (From Baxa and Bruhns 1967: 302; Blume 1985: 4; International Sugar Organization 1994: 285.)

Figure II.F.2.2. A Caribbean sugar factory. Although simplified, the drawing gives an idea of the activity. (From Pomet 1694: 94.)



pattern was established several hundred years ago (Figure II.F.2.2). Transport in leaky sailing ships meant that sugar inevitably ran the risk of contamination from sea water, and in the hot, humid conditions of a long voyage across tropical seas, sugar crystals coalesced. Thus, there was no point in making a finished product if it was only going to be damaged in transit to the market. In addition, the lack of fuel in the cane-growing regions meant that it was rarely possible to produce the finished product there. The industry does make its own fuel, bagasse, which is the residue of the stems after milling, but during the early centuries of the industry, inefficient furnaces meant that it had to be supplemented with another kind of fuel, usually scarce timber.

In short, the location of refineries in the importing countries reflected the realities of the sugar trade of yesterday. Today, there are clean, rapid, purpose-designed ships, so that long voyages pose little risk of damage to the sugar. And fuel is no longer the problem it once was. A long record of improvements in the design of furnaces and machinery means that an efficiently operated factory produces a surplus of bagasse. Nevertheless, factors continue to favor the location of the refineries in the importing, temperate countries.

Perhaps most importantly, refineries are not limited by harvests but can operate year-round by buying their supplies of raws from producers throughout the world according to season and price. Since the nineteenth century, refineries in some countries have also had the option of adapting to domestic beet sugar. By contrast, a refinery in a cane-growing region, dependent on the local production of raws, can only oper-

ate seasonally, which is not cost-efficient unless it is possible to stockpile or import raws to keep the refinery constantly in operation.

Because of such developments, many sugar factories now skirt the issue of refining by making direct-consumption sugars which, as their name implies, require no further processing. Whites, also known as plantation whites, are the most important of this class of sugar, and have a pol. of up to 99.5°. Whites were intended originally only for local consumption, but they have now entered international commerce to compete with refined sugars (Blackburn 1984: 290-3, 313-14).

Despite what might seem as the diminution of their locational advantages, however, the importing refineries will not yield their role easily. Vested interests continue to be involved, but additional refining is the only value-added part in the sugarcane industry that can be transferred from the producing to the importing countries in order to create jobs. Such refineries also have a trump card - they are close to and know their markets and can create demand as well as deliver what the market wants.

A glance along grocery store shelves reveals the many products of the refineries. There are the standard granular sugars, ranging from coarse to fine, and powdered sugar for icing cakes. The familiar sugar cubes for tea and coffee have been there since as long ago as the mid-1870s (Chalmin 1983: 84). The brown and yellow sugars, whether granular or in lumps, are still called Demerara - named after an administrative division of British Guiana, where, in the mid-nineteenth century, partially modernized factories deliberately produced sugar still colored and flavored by molasses.

It became a recognized trade name that could only be applied to this type of sugar produced in Demerara until 1913, when a court of law in London accepted the argument that the term had become common usage for any brown sugar, whether or not it had been made in Demerara (Barnes 1974: 11). Today, Demeraras are, in general, refined sugars that have been colored and flavored.

Other Products

Sugar is the most important product of the sugarcane industry. But there are also others, one being fuel. Because sucrose can be fermented, as well as crystallized, it is turned into an alcohol that substitutes for gasoline in cars. A substantial portion of Brazil's annual sugarcane crop is used for this purpose. Another use is the creation of fancy molasses (made from clarified, concentrated, very sweet syrup-like juice from which no sucrose has been drawn). This is what is used in cooking, spread on bread and scones, and employed as a topping for pancakes. Fancy molasses was a specialty of Barbados, with North America once the major market, where it sold for less than maple syrup. Barbados still exports fancy molasses, and a similar product is made in several other countries.

In addition, there are by-products of the manufacture of sugar. The most important are molasses, alcoholic beverages, bagasse, and filter mud, each of which has several uses. The molasses that spins from the massecuite in the centrifuges of modern factories is known as final or blackstrap molasses. It contains unextractable sucrose, as well as glucose, fructose, water, and minerals, the composition varying according to the climate and soil in which the cane was grown. Relatively little of this type of molasses enters international trade. One local use is for animal feed (either fed directly to the animals or mixed with other foods). Other derivatives of molasses include industrial alcohol, citric acid, and yeast. But perhaps its best-known use is in the manufacture of alcoholic beverages.

Rum is the sugar industry's best-known drink, flavored and colored with burnt sugar. But there are numerous other cane liquors, such as the *cachaça* of Brazil. Moreover, in cane-growing countries, alcohol is now mixed with imported concentrates to make gin and even whisky. "Choice" molasses, the by-product of noncentrifugal sugar, is also distilled for alcoholic beverages.

Bagasse – the dried cane pulp remaining after the juice is extracted – is a fortuitous by-product of sugar manufacturing that finds important uses. In some countries it is a fuel that generates electricity for the national grid. Ongoing research has also pointed the way to other applications, and it is now employed in the production of compressed fiberboard, various types of paper, and plastics.

Much of the filter mud, the result of clarification of

the cane juice, is returned to the cane fields as a fertilizer. But it is also possible to extract a crude wax from it that can be used in the manufacture of polishes. All of these by-products are major contributors to the profitability of the sugarcane industry (Barnes 1974: 446–69; Blackburn 1984: 314, 327–47).

Historical Geography²

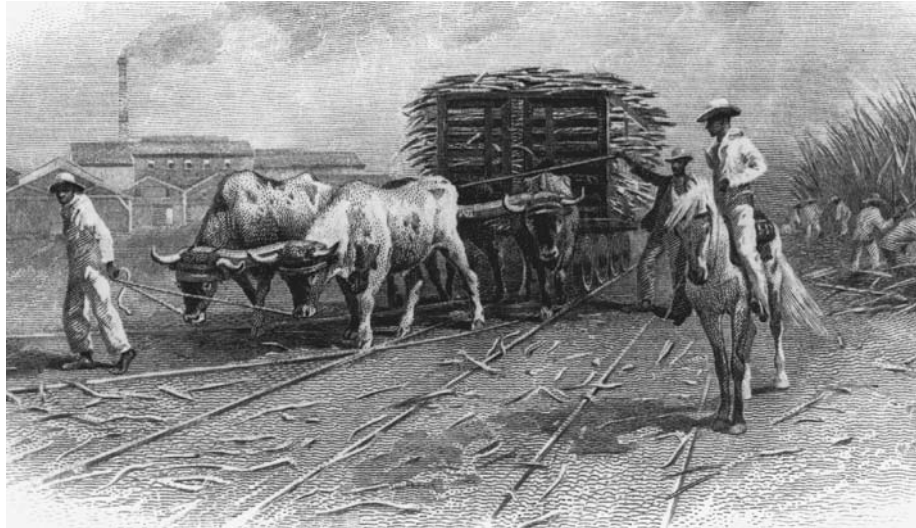
The manufacture of crystal sugar from sugarcane is one of the world's oldest industries. Its history is markedly episodic, with periods of technological innovation and geographical expansion separating periods of comparatively little change. Sugar became closely linked to major forces in European history when the Portuguese, who, in the fifteenth century, were venturing along the west African coast, discovered that the production and sale of sugar grown on nearby islands, especially Madeira, could be a means of financing the enterprise of reaching the East Indies. In subsequent centuries, sugar became intimately associated with the European colonization of the tropical portions of the New World (Figure II.F.2.3). The demand for labor that it created led to the transatlantic slave trade that carried millions of Africans to the Americas. These were followed by East Indians and other Asians in still more international migrations.

Yet sugar was hardly confined to the Americas, and neither were the historical forces that have shaped the industry from the rise of capitalism through the industrial revolution to twentieth-century advances in science and technology. The first use of sugarcane by humans was doubtless to chew it to obtain the sweet juice, a practice that continues, especially among children. Later, in southern Asia at some unrecorded time, efforts were begun to extract the juice with a simple mill or press and to concentrate it by boiling into a sweet, viscous mass. The first evidence of crystal sugar production appears at about 500 B.C. in Sanskrit texts that indicate it took place in northern India. They describe in rather vague terms the making of several types of sugar for which the principal use seems to have been medicinal. Knowledge of this technique spread from northern India eastward to China and (along with the cultivation of sugarcane) westward into Persia, eventually reaching the east coast of the Mediterranean about A.D. 600.

Sugar making in India and China, however, remained a technologically conservative, small-scale, largely peasant activity until the twentieth century. In China, for reasons that are not entirely clear, sugar joined, rather than displaced, its competitors, becoming one more option in a cuisine that continued to value honey, manna, maltose, and sagwire palm sugar. Clearly, the history of sugar in the East stands in marked contrast to that which unfolded in the West.³

The sugar industry entered the Mediterranean basin as part of an agricultural revolution carried out by the Arabs. In the years following the founding of

Figure II.F.2.3. Oxen drawing a cart of cane, with mill in background. (From a bond of the Eastern Cuba Sugar Corp., issued in 1922.)



Islam (Watson 1983), its adherents presided over the introduction of tropical and subtropical crops from Asia to the countries of the Mediterranean, as well as the techniques of irrigation that permitted cultivation throughout hot, dry summers.

To mill sugarcane, the burgeoning industry borrowed existing Mediterranean technology for extracting oil from olives and nuts and, in a second operation, used screw presses to obtain more juice from the bagasse. The juice was then clarified, reduced to the point of crystallization in open pans over furnaces, and the resulting syrup was placed in conical pots from which the molasses drained, leaving a loaf of sugar in each pot. The mills relied on water, animals, and men for power.

The risk of frost in the Mediterranean, together with the need for irrigation water, confined the cultivation of sugarcane to a few favored locations in the Levant, North Africa, Cyprus, Crete, Sicily, and Andalusia. Sugar production on the Atlantic islands of Portuguese Madeira and the Spanish Canaries was also Mediterranean in character. In its Mediterranean phase, however, the industry was labor-intensive, small-scale, and unable to produce large quantities of sugar, which therefore remained a luxury product. In short, there was little change in the technology of the Old World sugar industry between its introduction to the Mediterranean and its decline in the face of competition from the New World nearly 1,000 years later.

This Mediterranean phase did, nevertheless, foreshadow what was to come in the New World: Atlantic sugar was produced by slave labor brought from Africa to plantations in a system like that which would arise in the Americas. Moreover, the manner in which the Venetians and Genoese organized the supply of sugar to continental Europe during the Mediterranean phase also served as something of a blueprint for

things to come. The capital of these Italian city-states established a colonial relationship with the producing regions. They built the first refineries in Bologna and Venice and also drew on slave labor from the coasts of the Black Sea. The Genoese came to concentrate their investments increasingly in the western Mediterranean and were later instrumental in helping to finance the transfer of the industry to the Americas.

The years between 1450 and 1680 mark that transfer, as well as the decline of the Mediterranean industry in the face of American competition. It was a phase not only of technological innovation but also of geographical expansion, with the sugar-producing islands of Madeira, the Canaries, and São Tomé serving as stepping stones across the Atlantic. The Spanish first cultivated sugarcane on Hispaniola in the early 1500s, but the Portuguese, beginning in Brazil during the 1520s, were initially far more successful. The English turned Barbados into a sugar colony during the 1640s, and the sugar industry gradually spread across the Caribbean and to other regions of the New World.

In tropical America, not only did the climatic conditions favor sugarcane cultivation but so did the abundance of land to grow it on and the forests needed to provide fuel for the factories. With these advantages, the American sugar industry assumed an altogether new scale in terms of number of mills, size of land holdings, and quantity of exports. Such an increased scope of activity created a large demand for labor, which the Portuguese, at first, tried to satisfy by employing the indigenous population. But as that population declined in the face of Old World diseases, they turned increasingly to Africa, and the entire American industry soon depended on slave labor from that continent, resulting in the importation of more than 10 million people (Curtin 1969).

Still another development was a technological revolution in the seventeenth century that saw the replacement of the inefficient Mediterranean-style mills used initially in the New World (Galloway 1993: 211–22). Perhaps the most important innovation was a mill which, in its basic design, consisted of three vertically mounted rollers that could be turned by animal, water, or wind power. Such a mill did not require the stems to be chopped into pieces before milling, and the efficiency of its operation made the use of presses redundant. It is interesting to note the strong evidence of the contribution of Chinese technology to this design (Daniels and Daniels 1988). More efficient milling demanded, in turn, more efficient manufacture of sugar, and notable improvements in the design of furnaces allowed a production-line organization for transforming cane juice into crystal sugar.

Sugar colonies with slave-run plantations characterized American industry in the eighteenth century, a period once again of relatively little change in technology, although planters did gradually introduce conservationist measures to preserve the fertility of the soil and supplies of fuel. The sugar colonies exported sugar and rum to Europe and North America and imported slaves and food. Such an economy supported the densest concentrations of population in colonial America.

With the nineteenth century, however, there began another era of innovation in methods of production and of geographical expansion. The successful slave revolt in the French colony of St. Domingue (now Haiti) during the 1790s heralded the beginning of the end of the old regime. Nonetheless, the process of abolition was only completed a century later when the slaves of Cuba and Brazil finally gained their freedom. Many of the former slaves continued to work on sugar plantations, but planters also sought other sources of labor. Between 1838 and 1917, the British recruited workers from their Indian empire and, in so doing, profoundly changed the ethnic composition of many of their colonies. Laborers also came from Madeira, China, Japan, and the Pacific islands to work in the cane fields.

Major nineteenth-century technical innovations included steam power and the replacement of the three-roller mills by horizontally laid iron rollers mounted in sets of three that could be linked with other sets to form a train of six, nine, and even more rollers. With some improvements, this basic design of the mill has continued to the present. Cane-sugar technology also borrowed from the beet industry, especially the use of vacuum pans, in which juice could be boiled at lower temperatures under reduced pressure to save fuel. The centrifuges for separating the crystals from the molasses were also borrowed, and large central factories, which could accommodate the new machinery, gradually replaced traditional mills.

Still another major development was the breeding of new sugarcane varieties. The new technology, coupled with improved varieties of the raw material, helped keep the sugarcane industry competitive with the sugar-beet industry. By the end of the nineteenth century, the manufacture of cane sugar was becoming a capital-intensive agribusiness that relied on continued improvements in technology and methods of cultivation to keep a competitive edge.

During the nineteenth century, the link between the sugar industry and European expansion continued. A rising prosperity among the growing populations of western Europe led to increased demand at a time when European territorial empires extended into Africa, southern Asia, and the Pacific. One result was that sugarcane industries were established in such places as South Africa, Java, Queensland, Fiji, Hawaii, and Taiwan. In southern Asia, the Western, industrial style of sugar production either replaced or operated alongside traditional noncentrifugal producers. By 1900, the sugarcane industry had become one of the major economic activities of the entire tropical world.

The Sugar-Beet Industry

The Raw Material

Sugar beet is a root crop of temperate lands in Eurasia and America that in recent years has also become an important winter crop in North Africa and the Middle East. Obviously, it has adapted to a wide range of climatic and soil conditions, even growing well in the short summers of Finland, the dampness of Ireland, the high altitudes of Iran and Sichua, and the hot, dry Imperial Valley of California. It benefits from irrigation and from long hours of daylight. Differences in temperature, day length, and rainfall do, of course, influence the sugar content of the roots. It is a biennial, storing food in its swollen roots to carry the plants through the first winter and the process of setting seed in the second year. Farmers harvest the roots that contain the sugar at the end of the first year's growing season, thus interrupting the plant's natural cycle. Toward the polar limits of the sugar beet's range, low summer temperatures and extended length of the day encourage some plants to set seed prematurely during the first year, resulting in poor development of the roots. Plants that do this are known as "bolters," which if present in significant numbers, lower the yield of sugar from a field.

Cultivated and wild beets belong to the genus *Beta*, of the family Chenopodiaceae. The species *Beta vulgaris* L. includes the common beet, the mangel-wurzel, and the sugar beet. All three have descended from wild sea-beet, *Beta maritima*, by human selection. The Romans used beets, probably *Beta maritima*, as food for both humans and animals and thereby selected for its value as a vegetable. Cultivators in medieval and early modern Europe developed

the roots for animal feed, and since the eighteenth century, the capacity of *Beta vulgaris* to store sucrose in its roots has been the focus of breeding. Selection in Germany increased the sugar content of the roots from 7 percent in the eighteenth century to between 11 and 12 percent by the 1850s. The sugar content is now up to 20 percent. In addition to this success, researchers also breed sugar beets to discourage "bolting," to resist disease, and for shape and fibrosity to help in harvesting and milling. Sugar beets provide an example of a rapidly domesticated plant that is still being modified and improved to suit a particular purpose (Bailey 1949: 353; Campbell 1976; Bosemark 1993; Evans 1993: 101–3, 107, 109).

The main stages in the extraction of sucrose from beet have remained basically the same over the last century or so, but there have been improvements in the efficiency of each stage. On arrival at the factory, the beets are washed to remove soil and stones and then sliced into thin, smooth pieces known as "cossettes." The aim is to maximize the surface area of the beet so as to facilitate the diffusion of the sucrose. The cossettes then enter a countercurrent diffuser through which they move against the flow of a hot water extractant. This operation transforms about 98 percent of the sugar from the beet into a raw juice. The juice, in turn, is purified, reduced by evaporation, and crystallized, and the crystals are separated in centrifuges from the mother liquor. Formerly, beet sugar factories produced a raw sugar that was further treated in a refinery; now many factories produce a white sugar that is 99.9 percent pure (Vukov 1977: 13, 421, 426–7; Reinefeld 1979: 131–49; Bichsel 1988).

More than 90 percent of the revenue of the sugar-beet industry comes from sugar. Alcohol production is the best financial alternative to making sugar, but researchers have been unable to generate other by-products that are very lucrative. The lime sludge from the purification process is sold for fertilizer. Citric acid and baker's yeast are produced by fermenting the molasses, and the exhausted cossettes can be used for animal feed (Blackburn 1984: 338–9; Tjebbes 1988: 139–45).

Historical Geography

The sugar-beet industry is only two centuries old. In 1747, a Berlin professor of chemistry, Andreas Marggraf (1709–82), succeeded in extracting a modest quantity of sugar from beet. Although he published the results of his research in French and German, he did not put them to commercial use (Baxa and Bruhns 1967: 95–9). However, his student Franz Carl Achard (1753–1821) was more practical. He improved the raw material by breeding the cultivated fodder beets for sugar content, and he evolved the white Silesian beet, which is the ancestor of all subsequent sugar-beet varieties (Oltmann 1989: 90, 107).

In the years around 1800, Achard was active in promoting the beet industry, and in 1801, with the

financial assistance of the King of Prussia, he began to build what may have been the world's first sugar-beet factory (Baxa and Bruhns 1967: 113). Although it was not a financial success, other Prussians followed his initiative, building several small factories in Silesia and around Magdeburg. Russia was the second country to enter the industry: Its first sugar-beet factory opened in either 1801 or 1802 (Baxa and Bruhns 1967: 118; Munting 1984: 22), and the first factory in Austria opened in 1803. In the beginning, the French limited themselves to experiments, as did the Dutch, whose Society for the Encouragement of Agriculture offered a prize for extracting sugar from native plants (Slicher van Bath 1963: 276–7; Baxa and Bruhns 1967: 99–119).

These experimental and very small scale beginnings of the sugar-beet industry were given a considerable boost during the Napoleonic wars. In 1806, Napoleon's ban on the import of British goods and Britain's retaliatory blockade of his empire greatly reduced the supplies of cane sugar that reached continental Europe. Napoleon encouraged the production of beet sugar as a substitute, and landowners in France and the countries north of the Alps tried to respond.

The paucity of seed and unfamiliarity with the requirements of the crop led, however, to a disappointing supply of beet, part of which rotted on the way to the factory because of poor transportation. The number of factories illustrates the extent to which policy overreached reality: In France, in the season that spanned 1812 and 1813, only 158 factories of the 334 for which licenses had been given were actually in working order (Baxa and Bruhns 1967: 139). With the low yields of beet per hectare, low sucrose content, and a disappointing rate of recovery of the sucrose in the factories, beet sugar could not compete with cane once imports from the West Indies resumed after 1815. The beet industry disappeared from Europe except in France where it hung on until better times (Slicher van Bath 1963: 277; Baxa and Bruhns 1967: 134–45).

Those times began in the late 1830s and gathered force throughout the middle of the century, benefiting from improvements in both field and factory. In France, P. L. F. Levêque de Vilmorin (1816–60) was particularly successful in breeding varieties of beet for greater sugar content, with improvements continuing after his death (Baxa and Bruhns 1967: 190–1). In the first generation of sugar-beet factories, the juice was extracted by grinding the beet in animal-powered mills and then placing the pulp in presses (Baxa and Bruhns 1967: 148). In 1821, however, another Frenchman, Mathieu de Dombasle (1777–1843), proposed the procedure of slicing the beets and extracting the sucrose in a bath of water. He called the method maceration, but it is now known as the diffusion process. In 1860, Julius Robert (1826–88) became the first to employ it (Baxa and Bruhns 1967: 150, 176).

Diffusion replaced the mills and presses and remains to this day a standard part of the processing of sugar beet. Vacuum pans were first used in 1835 in a beet factory in Magdeburg, and centrifuges became part of factory equipment during the 1840s (Baxa and Bruhns 1967: 152, 172). An additional development encouraged the revival of the beet industry – the arrival of cheap Russian grain in western Europe where, from the 1820s onward, it caused a fall in grain prices. Western European farmers who had been growing grain now required a substitute crop, and beet was a good candidate. It could fit into the agricultural rotation, growing on land that would previously have been left fallow, and its leaves and roots provided feed for animals. This, in turn, made possible an increase in livestock numbers, which meant more manure. If the roots were sold to a factory, they earned the farmer cash, and the pulp could also be used for feed (Galloway 1989: 131).

Despite the advantages of sugar beet to the agricultural economy and improvements in its raw material as well as factory technology, the beet industry nevertheless was still not competitive with cane. Rather, its revival in the 1830s, and its continued growth, depended on government protection through tariffs on imported cane sugar and incentives of one sort or another, such as subsidized exports. The 1902 Brussels Convention attempted to bring some order to a scene characterized by a protected beet industry and a complaining sugarcane industry, yet protection for beet sugar remains in place (Chalmin 1984: 9–19; Munting 1984: 21–8; Perkins, 1984: 31–45).

The revival of sugar-beet cultivation began in northern France, where it had never entirely died out, and continued in Prussia and the other German states during the 1830s, in Austria-Hungary in the 1840s, and in Russia in the 1850s. By the 1850s, Germany had become the most important producer of beet sugar, responsible by the end of the century for rather more than a third of the European total. By this time, beet cultivation extended from southern Spain in a curve arching north and east through France, the Low Countries, Germany, and eastern Europe, and into the Balkans, Russia, and the Ukraine. It also extended into Denmark and southern Sweden. The industry was particularly important in northern France, in the Low Countries, around Magdeburg in Germany, in Bohemia, and around Kiev in the Ukraine. Great Britain was noticeably absent, refusing to subsidize a beet industry of its own, but rather buying sugar, whether cane or beet, wherever it was cheapest.

Sugar beet remained a predominantly European crop throughout the twentieth century. In the years approaching 1990, Europe accounted for about 80 percent of the world's production of beet sugar, with 40 percent of that production coming from the European Union. Since 1991, production in the countries of the former Soviet Union has lost ground, but this is

probably a temporary situation. The geography of the European beet-sugar industry has also remained remarkably constant: Those regions that were important producers at the beginning of the twentieth century remained so at its end. There has been some modest expansion on the periphery into Ireland and Finland, and since the 1920s, Great Britain has finally developed a sugar-beet industry of its own. Two considerations led to the change in Britain's policy towards beet: World War I had revealed the difficulties of relying heavily on continental European producers of sugar and the awkwardness of such dependence. Moreover, sugar beet had the potential of being a useful cash crop for farmers in the agricultural depression that followed the war.

The North American sugar-beet industry dates from the 1880s and has increased steadily, producing nearly 4 million tonnes of sugar a year. The industry is overwhelmingly located in the United States. Beet is grown in the Midwest, but the bulk of the crop grows on irrigated land in the West. In Asia the industry became significant only in the 1920s and has seen rapid expansion since the 1960s (Baxa and Bruhns 1967: 192–201, 221–2, 262–94; International Sugar Organization 1994: 279–85).

Like the cane industry, the sugar-beet industry invests heavily in research. The breeding of new varieties, the methods of harvesting and planting, and improvements in factory technology are all important foci of attention.

The Contemporary Sugar Industry⁴

Production

At the beginning of the twentieth century, beet-sugar production exceeded that of cane sugar, with the total combined production of centrifugal sugar about 12 million metric tonnes raw value (mtrv). However, today cane accounts for most of the sugar produced (about two-thirds). Total combined production reached 100 million mtrv in the mid-1980s and rose to 120 million mtrv by the mid-1990s. This expansion has been fueled by increases in consumption of about 2 percent a year, resulting from a growing world population and improving standards of living in some of the less-developed countries of the world. Non-centrifugal sugar also continues to be an important sweetener: Statistics are almost certainly incomplete but show production in excess of 15 millions tonnes at present.

The sugar industry in some countries is highly protected. The United States, for example, maintains a domestic price for sugar well above the world price and controls the amount of sugar imported. The European Union protects its beet growers, and the industry in India is carefully regulated to control production and domestic prices. Clearly, the interventionist traditions in the industry established long ago remain very much alive.

India is the world's major producer of cane sugar, and its sugar industry continues to grow. Annual centrifugal production has reached 16 million mtrv, which includes nearly 1 million tonnes of khandsari sugar. India is also the world's major producer of non-centrifugal sugar, accounting for perhaps as much as two-thirds of the total. Practically all of this sugar is consumed in India; only rarely, after exceptionally good harvests, are small quantities exported.

Brazil's production has increased rapidly in recent years to reach 13 million mtrv, plus some small-scale production of noncentrifugal sugar known as *rapadura*. The country has the advantages of abundant land and a good climate, and its production is divided between the home sweetener market, fuel alcohol for automobiles, and exports. Brazil has the ability to direct sugar to exports or to fuel alcohol, depending on the world prices. Cuba and Thailand compete for the third and fourth rankings among the world's sugar producers. Cuba's annual production in the late 1980s was around 8 million mtrv, but collapsed to half this amount when the fall of communism in Eastern Europe brought about a loss of its major markets. The Cuban industry's ability to recover remains a major question. Thailand has only recently become an important sugar producer. Its industry is expanding, and production is now in excess of 6 million mtrv, of which two-thirds is exported.

The European Union is responsible for nearly half of the world's beet sugar, about 18 million mtrv annually. Germany and France are the main producers, although Ukraine and Poland produce close to 4 million and 2 million mtrv, respectively. By and large, the beet industry in eastern Europe suffers from poor management and a lack of investment in machinery, and because much land is still publicly owned. The region has enormous potential, however; several western European companies have taken advantage of privatization schemes to buy factories and begin the work of modernization. The United States, China, and Turkey are also major beet-sugar producers. The United States and China are among the relatively small number of countries which, because they extend across a wide range of climatic zones, are able to grow both cane and beet.

Trade

International trade in centrifugal sugar amounts to about 30 million mtrv, or about one-quarter of the total world production, meaning that most sugar is consumed in the country where it is produced. Much of the trade takes place under special arrangements, and only a small portion of the sugar traded internationally sells at the free-market price.

The European Union buys sugar at a preferential price from the former colonies of its member states; the United States allocates import quotas to a large number of countries. Cuba bartered sugar for oil with the former Soviet Union, and now barter with Russia

on a more limited scale. These arrangements have had what might seem curious consequences. The European Union is both a major importer (1.8 million mtrv annually) of sugar from its former colonies and a major exporter (five million mtrv annually) because of its beet production. Some countries (Barbados, Jamaica, Guyana, and the Dominican Republic) export all, or nearly all, of their sugar at a premium price to meet contractual arrangements with the United States and the European Union, and they import at world market price for their own consumption. Refineries also import raw sugar, only to export it again in a practice known to the trade as tolling. This accounts for the presence of the United States on the list of sugar exporters. Quotas limit its imports, but tolling permits the use of otherwise surplus refining capacity and provides employment.

About 75 countries export sugar and about 130 import it – the numbers fluctuate a little from year to year. Most trade in small amounts, and many are minimal participants dealing in less than 10,000 tonnes a year. But the European Union, Ukraine, Cuba, Brazil, Thailand, and Australia all export more than 1 million mtrv annually. Together, these latter countries account for by far the greater part of sugar exports. By way of contrast, the European Union, Russia, Canada, the United States, Japan, and South Korea all import more than 1 million mtrv a year apiece, with Malaysia and Algeria not far behind. Such activities provide certainties, insofar as there can be any, in the sugar trade. The major sources of uncertainty are India and China. They may appear unexpectedly on the market as either importers or exporters if their policies change or if their policy makers misjudge the market situation. Weather is also an uncertainty. Poor or excellent harvests in a number of countries can cause shortages or create surpluses.

In most countries there are stocks of sugar held against sudden shortages and increases in price. The world stocks-to-use ratio in the mid-1990s was considered high, at about 19 percent (USDA 1996: 9). This translates into low free market prices because there are reserves to draw on in case of a sudden shortage. Sugar traders and some governments carefully monitor production, import, export, and consumption data in each country and calculate the buildup and/or drawdown of stocks with a view to predicting demand and prices. There is a very active trade in sugar futures.

Competition

For several hundred years, sucrose in the Western world has been without a serious competitor in the sweetener market, but recently this has changed. The leading caloric competitor is high-fructose corn syrup (HFCS). It is a liquid sweetener made from plants (especially maize) that contain a sufficient amount of starch, although sweeteners are also made from sweet potatoes and tapioca in Asia and from wheat and

potatoes in Europe. HFCS appeared in the 1970s, during a period of high sugar prices, but continued in production after sugar prices fell. Sugar usually has a price advantage over HFCS, and HFCS is not always a good substitute for sugar. Bakers and manufacturers of confectionery and cereals prefer sugar because of its "bulking, texture and browning characteristics" (USDA 1995: 18). HFCS is most competitive in the manufacture of soft drinks. The United States is by far the largest producer of HFCS, with the European Union, Canada, Korea, Argentina, and Taiwan making very modest quantities. HFCS has captured rather less than 10 percent of the sweetener market and, in the immediate future, is not expected to expand beyond the liquid sweetener market in a few countries (USDA 1995: 15–20).

Low-calorie, high-intensity sweeteners have gained in significance since the 1980s and claim a small percentage of the sweetener market. Saccharin and aspartame are perhaps the best known. They are attractive to consumers concerned with diet and can compete in price with sugar. Both are used to sweeten coffee, tea, and other beverages ("table-top use"), but aspartame, benefiting from the demand for diet soft drinks, has received approval for a wider range of uses in the European Union, the United States, Canada, and Japan. This branch of the sweetener market is still evolving. Low-calorie sweeteners are used together in some applications, but they also compete with one another and, of course, they compete in the soft drink market with HFCS and sugar (Earley 1989; USDA 1995: 23–4).

The fact that the manufacture of sugar is one of the oldest industries in the world gives the sugar industry a special place in the cultural history of food, but other characteristics also give it a particular interest. It is unique among agro-industries in that it has both tropical and temperate sources of supply of its raw material. It has been responsive throughout its history to developments in science and technology and today continues to invest in research. Government intervention is a long-standing tradition. The importance of sugar to the finances of European imperialism was a great incentive for the colonial powers to attempt to manage the international trade in sugar for their own benefit. Governments continue to intervene to this day with subsidies and protective tariffs to defend vested interests.

The sugarcane industry has had profound economic and social consequences on those parts of the world in which it became a major crop. Indeed, perhaps no other crop has had such a formative influence on societies. The industry has divided societies, along class lines, between the owners of the factories (either local elites or – often nowadays – foreign companies) and the work force. Even where the factories have been nationalized, or are owned by cooperatives, disparities exist in income and prestige between managers and laborers. Because of its great demand for labor – satisfied first by African slavery, then by

indentured workers, and finally by free laborers – the industry also produced multiethnic populations that frequently have complex internal politics. Another legacy is economic dependency. Although the empires are gone, many ex-colonies continue to grow sugar as a staple, relying on special arrangements with their former rulers (now members of the European Union), as well as with the United States, that enable them to sell their sugar at a premium above the low price that sugar generally commands on the open market.

The dilemma of dependency has had no easy resolution. Sugarcane producers have little bargaining power, given the oversupply of their product, and alternatives to the cultivation of sugarcane that can provide a better level of employment and income have been difficult to find. Dependency is keenly felt by the populations of the cane-growing countries, and where this sense of frustration is joined with the memory of slavery, as in the Caribbean, sugarcane is a very problematic crop.

J. H. Galloway

Notes

1. This statement requires one reservation. The medieval Mediterranean sugar industry and the early American industry cultivated only one variety of cane, known now in the literature as Creole cane. Opinion differs on the correct botanical identification of this cane. Most authorities accept that it was either a variety of *S. officinarum* (Blackburn 1984: 2, 91) or a hybrid of *S. barberi* and *S. officinarum* (Stevenson 1965: 41). However, Fauconnier (1993: 1) considers it to have been a variety of *S. barberi*.
2. In this section I draw heavily on Galloway 1989.
3. The authoritative discussion of sugar-making in China is that by Christian Daniels (1996).
4. I have used three sources for the statistics in this section: Czarnikow (1997: 55; the International Sugar Organization (1994); and the USDA (1997: 31–6). There are some differences in these data. For a brief discussion of the problems of collecting statistics on the sugar trade, see Ahlfeld and Hagelberg (1989: 59–65).

Bibliography

- Ahlfeld, Helmut, and G. B. Hagelberg. 1989. Statistical problems in world sugar balance calculations. In *Sugar. Essays to mark the 125th anniversary of F. O. Licht*. Third edition, ed. Helmut Ahlfeld, 59–65. Ratzeburg, Germany.
- Bailey, L. H. 1949. *Manual of cultivated plants*. Rev. edition. New York.
- Barnes, A. C. 1974. *The sugar cane*. Second edition. Aylesbury, England.
- Baxa, Jacob, and Guntwin Bruhns. 1967. *Zucker im Leben der Völker. Eine Kultur- und Wirtschaftsgeschichte*. Berlin.

- Bichsel, S. E. 1988. An overview of the U.S. beet sugar industry. In *Chemistry and processing of sugarbeet and sugarcane*, ed. M. A. Clarke and M. A. Godshall, 1-7. Amsterdam.
- Blackburn, F. 1984. *Sugar-cane*. London and New York.
- Blume, Helmut. 1985. *Geography of sugar cane*. Berlin.
- Bosemark, Nils Olof. 1993. Sugar beet. In *Traditional crop breeding practices: A historical review to serve as a baseline for assessing the role of modern biotechnology*, ed. Organization for Economic Co-operation and Development, 123-36. Paris.
- Campbell, G. K. G. 1976. Sugar beet. In *Evolution of crop plants*, ed. N. W. Simons, 25-8. London and New York.
- Chalmin, Ph. G. 1984. The important trends in sugar diplomacy before 1914. In *Crisis and change in the international sugar economy 1860-1914*, ed. Bill Albert and Adrian Graves, 9-19. Norwich and Edinburgh, U.K.
- Chalmin, Phillippe. 1983. *Tate and Lyle, geants du sucre*. Paris.
- Chen, James C. P. 1985. *Meade-Chen cane sugar handbook*. Eleventh edition. New York.
- Clark, W. 1823. *Ten views of the island of Antigua*. London.
- Curtin, Philip D. 1969. *The Atlantic slave trade. A census*. Madison, Wis.
- Czarnikow and Company. 1997 *The Czarnikow sugar review*, No. 1883: 95-6.
- Daniels, Christian. 1996. Agro-industries: Sugarcane technology. In *Science and civilization in China*, Vol. 3, ed. Joseph Needham, i-xxvii, 1-539. Cambridge.
- Daniels, John, and Christian Daniels. 1988. The origin of the sugarcane roller mill. *Technology and Culture* 29: 493-535.
1993. Sugarcane in prehistory. *Archaeology in Oceania* 28: 1-7.
- Earley, Thomas C. 1989. The effect of alternative sweeteners on the structure of the world sugar economy. In *Sugar. Essays to mark the 125th anniversary of F. O. Licht*. Third edition, ed. Helmut Ahlfeld, 191-4. Ratzeburg, Germany.
- Evans, L. T. 1993. *Crop evolution, adaption and yield*. Cambridge.
- Fauconnier, R. 1993. *Sugar cane*. London.
- Galloway, J. H. 1989. *The sugar cane industry. An historical geography from its origins to 1914*. Cambridge.
1993. The technological revolution in the sugar cane industry during the seventeenth century. In *Producción y comercio del azúcar de caña en época preindustrial*, ed. Antonio Malpica, 211-22. Motril, Spain.
1996. Botany in the service of empire: The Barbados cane-breeding program and the revival of the Caribbean sugar industry, 1880s-1930s. *Annals of the Association of American Geographers* 86: 682-706.
- International Sugar Organization. 1994. *Sugar year book*. London.
- Munting, Roger. 1984. The state and the beet sugar industry in Russia before 1914. In *Crisis and change in the international sugar economy 1860-1914*, ed. Bill Albert and Adrian Graves, 21-8. Norwich and Edinburgh, U.K.
- Oltmann, Wilhelm. 1989. Die Züchtung von Zuckerrüben - Ihr Beitrag zur Verbesserung von Quantität und Qualität. In *Sugar. Essays to mark the 125th anniversary of F. O. Licht*. Third edition, ed. Helmut Ahlfeld, 87-109. Ratzeburg, Germany.
- Pack, Christopher, ed. 1996. *The Czarnikow sugar review*, No. 1868.
1997. *The Czarnikow sugar review*, No. 1883.
- Perkins, John. 1984. The political economy of sugar beet in imperial Germany. In *Crisis and change in the international sugar economy 1860-1914*, ed. Bill Albert and Adrian Graves, 31-45. Norwich and Edinburgh, U.K.
- Pomet, Pierre. 1694. *Histoire générale des drogues*. Paris.
- Reinefeld, E. 1979. Progress in the technology of beet-sugar. In *Sugar: science and technology*, ed. G. G. Birch and K. J. Parker, 131-49. London.
- Slicher van Bath, B. H. 1963. *The agrarian history of western Europe A.D. 500-1850*. London.
- Stevenson, G. C. 1965. *Genetics and breeding of sugar cane*. London.
- Tjebbes, Jan. 1988. Utilization of fiber and other non-sugar products from sugarbeet. In *Chemistry and processing of sugarbeet and sugarcane*, ed. M. A. Clarke and M. A. Godshall, 139-45. Amsterdam.
- USDA (U.S. Department of Agriculture). 1995. *Sugar and sweetener. Situation and outlook report*. December issue.
1996. *Sugar and sweetener. Situation and outlook report*. June issue.
1997. *Sugar and sweetener. Situation and outlook report*. December issue.
- Vukov, Konstantin. 1977. *Physics and chemistry of sugar-beet in sugar manufacture*. Amsterdam.
- Watson, Andrew M. 1983. *Agricultural innovation in the early Islamic world*. Cambridge.



II.G

Important Foods from Animal Sources

II.G.1 American Bison

The American bison (*Bison bison*) is more closely related to cattle than to true buffalo, such as the water buffalo (*Bubalus bubalis*). Nonetheless, early European settlers called the unfamiliar animal they encountered in North America a “buffelo” [*sic*] and this misnomer has persisted to the present. Because we are so accustomed to thinking of the American bison as a buffalo, the terms are used interchangeably in this chapter.

Long perceived to be an environmental casualty of the conquest of the Great Plains, the American bison has gradually reasserted its presence in North America. Today, there are around 200,000 of the animals alive on the continent, and the danger of species extinction appears to have passed (Callenbach 1996). However, any further expansion of the population is likely to be linked, at least in part, to the animal’s economic usefulness, especially as a food source. At present, only a limited number of people are acquainted with the taste of bison. For buffalo meat to become part of the national diet, the advertising and agricultural industries will have to reintroduce a food which, at an earlier time, was essential to many of the continent’s inhabitants.

In size and appearance, the bison is imposing and distinctive. A mature male stands from 5 to 6 feet tall at the shoulder and may weigh from 1,800 to 2,400 pounds. Noticeably smaller, females seldom weigh more than 800 pounds. Despite its bulk, the bison is quick and agile and can sprint at speeds of up to 30 miles per hour. Like domestic cattle and sheep, it is cloven hooved; unlike them, both the male and female possess large, curved horns and a prominent hump at the shoulder. Buffalo are usually dark brown in color, although their hue lightens to brownish-yellow during the spring. The animals have two types of hair: a long, coarse, shaggy growth covering the neck, head, and shoulders, and a shorter, woolly growth found on the remainder of the body. During the spring, bison

shed much of this hair, assisting the natural process by rubbing against rocks or trees. Only the head, hump, and forelegs retain hair throughout the year.

In the past, wild bison migrated with the seasons, from north to south and back. Although their diet consisted of various grasses (blue stem, grama, and bunch), it was buffalo grass (*Buchloe dactyloides*) that served as their primary staple. A short, fine plant, buffalo grass requires little water, tolerates extreme temperatures, and is resilient in the face of grazing and trampling. As the bison population declined, this species became scarce as well. Competing plant life that was no longer destroyed by the bison’s hooves expanded throughout the plains, and without the wanderings of the bison to spread its seeds (either through manure or by attachment to the animal), buffalo grass was largely replaced by other grasses (Garretson 1938).

The buffalo’s mating or “rutting” season commences in early June, reaches a peak in late July and early August, and gradually concludes by the end of September. There is violent competition among the bulls to impregnate the females, followed by a nine-month gestation period and the birth of a single calf (McHugh 1972). In general, the calves mature quickly and are able to mate by their second year, although the animals are not fully grown until the age of six. On the average, the bison’s life span is between 20 and 30 years.

History of the Bison

When the European explorers first visited North America, large numbers of bison were present in perhaps 70 percent of the present-day continental United States, as well as in what are now the Canadian provinces of Alberta and Saskatchewan. Indeed, the pre-Columbian buffalo population is often estimated at some 60 million members. As a rule, the animals lived in groups of 20 to 40, gathering into larger herds only for rutting or migration.

Native Americans long utilized the bison as a food

source; archaeological evidence suggests that buffalo hunting was practiced more than 10,000 years ago. However, for the peoples residing in the vast plains stretching from the Rocky Mountains to the Missouri River, the buffalo represented more than simply a source of food and clothing.

In a physical sense, the animal provided the plains inhabitants with many items essential to survival in this environment, in the form of blood, meat, hide, bone, sinew, and manure, that were used for food, rope, weapons, shelter, blankets, clothing, fuel, and medicine (Dary 1974). In a spiritual sense, the buffalo provided the first Americans with still more. Understanding the animal to be one of numerous earthly representatives of a higher power, the Native Americans honored it as a spirit that influenced fecundity, happiness, strength, protection, and healing. Within such tribes as the Assiniboin, Cheyenne, and Kiowa, buffalo skulls served as central totems in ceremonial activities. Particularly revered was the rare albino, or "White Buffalo," believed to be the sacred leader of all herds. Among the Plains Indians, an understanding of the buffalo was perceived as essential for understanding one's own place within the universe (McHugh 1972).

Methods of preparing bison as food were largely determined by the gender of the cooks. When eaten fresh, the meat was often cooked by the hunters, whereas meat curing and preservation were tasks reserved for women. Hunting parties would sometimes use the hide as a cauldron in which to boil the meat. Variations on this practice included lining a hole in the ground with the animal skin or suspending the hide aboveground on sticks over a fire. In addition to boiling, the meat was frequently roasted by rotating it over an open fire.

To cure buffalo meat, Indian women relied on the sun, rather than on salting or smoking – the European methods. Selecting the choicest parts, the women cut the meat into strips, across the grain, in order to maintain alternating layers of lean and fat. These strips were then suspended on elevated racks in full sunlight for several days. The result was a jerky that could be eaten in the dried form or rehydrated by lengthy boiling. When cured, the meat was lightweight and largely imperishable, an ideal staple for a mobile culture (Catlin 1842).

The jerky, however, could be even further condensed when transformed into pemmican, a food whose name is a word from the Cree language – *pimikân*, derived from *pimii*, meaning grease or fat. Pemmican is aptly named. To make it, the Indians placed pulverized jerky into sacks sewn from buffalo hide and then poured hot, liquid marrowfat over it. The fat would both combine with the dried meat and serve to seal the sack. While still molten, pemmican could be formed into shapes convenient for storage or transport. Mostly produced in the form of 90-pound blocks, it was a staple for the Plains Indians, as

well as for the fur trappers and traders who spearheaded the Europeans' advance into the region.

As the Native Americans came to recognize pemmican's value in trade with the whites, additional ingredients were added to appeal to the latter's tastes. Although chokeberries became the most common supplement, some pemmican was flavored with cereals, sugar, and fruits. For trappers, this food was "the bread of the wilderness," a practical staple in terms of availability, stability, and nutrition. Whether the bison population was diminished as a result of such commerce is difficult to establish, although it is certainly probable (Binkerd, Kolari, and Tracy 1977).

Native Americans also incorporated buffalo into their diet in several ways other than as fresh and preserved meat. Various tribes developed methods of preparing blood soups and puddings. Roasted thigh bones were a popular source of tasty marrow, and the tongue was savored. Northern tribes enjoyed boiled fetal calves, and the Hidatsa specialized in grilled buffalo udders filled with milk. The Plains tribes occasionally dined on baked buffalo hides. However, this latter practice was infrequent as hides were generally more valuable for things other than food (McHugh 1972).

The Bison East of the Mississippi

Following the example of the first Americans, early European visitors utilized the buffalo as a source of steaks, stews, and marrow. Throughout the sixteenth century, Spanish explorers, such as Alvar Nuñez Cabeza de Vaca and Vincente de Zaldivar, wrote of consuming buffalo flesh during their adventures. The British, however, who settled east of the buffalo grounds, did not encounter the animal until they began to move inland at the beginning of the seventeenth century. By the end of the colonial era, the buffalo had virtually disappeared from the landscape east of the Appalachians.

The reasons for this disappearance are varied. Although settlers often hunted bison for food, it is doubtful that the human population of the time was numerous enough to eliminate great numbers of the animals. Rather, it seems more likely that as the whites transformed the landscape to meet their needs, the buffalo simply migrated westward away from people. In addition, as farmers began to raise more cattle, their grazing would have reduced the available food supply. Finally, it should be noted that the eastern herd was by no means as large as that of the west.

When European settlement expanded into the area between the Appalachians and the Mississippi, buffalo were once again on hand as a food source. In detailing his travels through Kentucky, Daniel Boone mentioned meals of fresh buffalo, and as other early frontiersmen penetrated the region, they often followed "buffalo roads" – narrow paths formed by herds in search of salt (Rorabacher 1970).

As had happened east of the Appalachians, however, as human settlement continued, buffalo east of the Mississippi became displaced by a changing economy. Initially, settlers relied heavily on bison for food; yet, once farms were established, the animals became nuisances, grazing on land more profitably used to raise cattle or crops. By 1830, the buffalo had disappeared from the East, and the American bison population was compressed into the Great Plains.

The Bison West of the Mississippi

Crowding into the plains created a variety of natural hazards for the buffalo. Food shortages were more common, and fatal diseases (particularly anthrax and bovine tuberculosis) could spread quickly and kill greater numbers as the herds grew more closely together (McHugh 1972).

Such consequences of the bison's westward migration, however, seem insignificant when compared with the destruction wrought by humans during the final third of the nineteenth century. It was then that the forces of economics and politics, as well as human greed, carelessness, and malice, came together in such a manner as almost to remove the American bison from the North American ecosystem. Between 1870 and 1873, the large buffalo herds of Kansas were destroyed. Next to go were the southern herds, and by 1878, the animals were absent from Texas to the edge of the prairie. The vast northern herds of the Dakotas and Montana survived a few years longer, perhaps into 1883. But by 1900, the herds of the Canadian plains were gone, and only about 300 of the animals remained in the United States.

Commercial hunting seems to have been the factor most significant in the buffalo's destruction. The earliest professional hunters in the West sought beavers, not bison, and required the latter only for food. By the early 1820s, however, commercial buffalo hunting had begun in southern Canada, with the primary aim of securing buffalo tongues (a popular delicacy) and hides for shipment to the East and abroad. Often, entrepreneurs recruited Native Americans to do such hunting in exchange for whiskey.

At least at this point, however, the bison's habit of shedding offered a degree of protection against excessive hunting. Because the sale of furry hides (for use as robes) produced the greatest revenue, and because the animals lost their furriness during the summer, hunting them was strictly seasonal. Nevertheless, the desire to trade for the white man's goods resulted in a Native American slaughter of buffalo that has been estimated as 30 percent greater annually than their personal needs required (Dobak 1996).

During the two decades prior to the Civil War, the Great Plains experienced a series of ecological changes that also hastened the disappearance of the bison. Foremost was the proliferation of open-range

cattle ranching. Indeed, by 1850, the prairie had become home to more than 50 million cattle that were major competitors for the region's food supply. Moreover, ranchers were unhappy with the seasonal wanderings of the bison, arguing that their grazing rendered the land unfit for beef cattle for as long as two years afterward (Vestal 1952).

At the same time, as human settlement increased and transportation on the prairie improved, the slaughter of the buffalo for industrial purposes escalated. The trade in buffalo robes accelerated as more forts and trading posts dotted the western landscape: In 1850, around 100,000 robes reached the markets of St. Louis, a record total at that time. The increase in civilian and military populations on the plains also meant that still more of the animals were slaughtered for food (Dary 1974).

After 1865, the future of the bison became ever more precarious. With the conclusion of the American Civil War, both the U.S. Army and the railroad corporations sought to consolidate their positions on the plains. Troops were stationed in the West to pacify the Native Americans, and thousands of men migrated to the prairie to lay tracks for the Union Pacific and the Topeka & Santa Fe railroads. Fresh buffalo meat served as the primary fare for both the soldiers and the laborers. Moreover, once the railroads began service through the West, the buffalo again (as they had in the East) became nuisances, routinely wandering onto or across the tracks, blocking trains, and causing long delays. George Bird Grinnell recalled two such delays while traveling on the Kansas Pacific, one of which lasted for around three hours (Grinnell 1892).

In 1870, a refinement in tanning techniques became still another factor in the decrease of the American bison. The animal's summer hide had long been dismissed as economically worthless because there was no known method for transforming the skin into commercial leather. However, once Philadelphia tanners devised such a method, the bison became a year-round target, and the promise of easy money encouraged more and more men to become buffalo hunters (Cronon 1991).

This was especially the case following the financial panic of 1873, when the nation entered a five-year period of severe economic depression. As banks closed and millions of easterners were suddenly unemployed, the seemingly inexhaustible bison herds of the plains offered a chance to earn a living, and perhaps even to acquire wealth. Buffalo hides brought from 2 to 3 dollars each during the 1870s, and a successful hunter might kill more than 250 animals in a single day. In the 1950s, Frank H. Mayer, once a "buffalo runner" himself, recalled:

The whole Western country went buffalo-wild. It was like a gold rush or a uranium rush. Men left jobs, businesses, wives, and children and

future prospects to get into buffalo running. There were uncoun­ted millions of the beasts – hun­dreds of millions, we forced ourselves to believe. And all we had to do was take these hides from their wearers. It was a harvest. We were the harvesters (Mayer and Roth 1958: 21).

Much of the time, hunters worked in gangs, with individuals specifically assigned to shoot or skin the animals. Larger parties would utilize horses to remove the hides quickly: After the buffalo's skin was loosened by making a series of cuts, it was tied to a horse which pulled the hide off the carcass (Wheeler 1923). The skins were awkward and bulky, often weighing more than 100 pounds each. Still, an experienced skinner could have the hide off an animal in about five minutes. Although the financial rewards encouraged efficiency, the animals' abundance more often produced carelessness and waste, and bison slain by commercial hunters were not really "harvested" at all but merely left to rot in the sun.

In general, the bison was easy prey. Seasoned hunters understood its habits and instincts and developed techniques to attack it when confused and vulnerable. Recalling the hunts of the 1870s, one participant stated that "[t]he stupidity of the buffalo was remarkable" (Wheeler 1923: 81). Although there was a modicum of risk in buffalo hunting, accidents were more often the result of human error than of aggressiveness on the part of the hunted. Many retired hunters later wrote about injuries caused by falling off horses or by rifles exploding. Occasionally, too, poor judgment during winter months resulted in death from exposure (Dodge 1959).

In addition to destruction by commercial hunters, the buffalo population was further decimated by increasingly numerous "sport-hunters." As early as the 1830s, such traveling gentlemen as Washington Irving began to test their marksmanship on the buffalo, and during the 1850s, an Irish nobleman, Sir George Goe, indulged in a highly publicized "safari" with 21 wagons and a staff of 40 servants. In a three-year period, his party killed 2,000 bison.

Such "sporting" expeditions became increasingly widespread as the century progressed, especially during the depression that ensued after the "Panic of 1873," when the railroads were determined to increase revenue by increasing the number of passengers going west. Rail executives hit upon the notion of catering to the sportsmen (who seemed to have money and time when most people did not) and advertised special "buffalo trains" that would transport easterners into areas populated by the animals. Upon arrival, the passengers could shoot bison from the comfort (and safety) of the railroad cars. More intrepid hunters might track them on horseback, although in general the procedure was to hire an experienced "scout" to minimize the danger.

The most renowned of such guides was William F. Cody, or "Buffalo Bill," a clever entrepreneur who later prospered as the creator of the traveling "Buffalo Bill's Wild West" show, on the road from 1883 to 1916. Much of Cody's initial fame had resulted from escorting prominent personalities, such as Gordon Bennett II (publisher of the *New York Herald*) and the son of Russian Tsar Alexander II (Aleksandre Nikolayevich), on hunts.

Such expeditions were bloody slaughters. Many hunters recalled ruining expensive rifles because of lengthy periods of continuous and rapid firing that caused them to overheat. If the hunters were successful in making a "stand" – a technique of tricking the buffalo into immobility – the potential for killing was limited only by their ammunition. Numerous parties recorded killing up to 200 animals per hunter per day.

The Final Buffalo Hunts

The impact of such slaughter was not unnoticed at the time. In 1871 and 1872, the territorial legislatures of Wyoming and Montana passed laws intended to reduce buffalo hunting. In the U.S. Senate (also during 1872), Cornelius Cole of California and Henry Wilson of Massachusetts attempted on separate occasions to enact federal protection for the bison. Both, however, were unsuccessful.

Two years later, U.S. Representative Greenburg L. Fort of Illinois sponsored national legislation prohibiting the killing of female bison by anyone not a Native American (Dary 1974). This bill triggered much debate that clearly indicated that its opponents saw in the extermination of the bison the most effective way of subduing the Plains Indians. Indeed, Secretary of the Interior Columbus Delano had endorsed such a position (in his annual report for 1873) by predicting that the disappearance of the bison would eliminate future Native American uprisings by reducing the Native Americans. One of the opponents of Fort's legislation, Samuel S. Cox, bluntly argued that the bill "favored Indians at the expense of white men," whereas only the reverse was desirable (Dary 1974: 127).

Despite strong opposition, the bill passed the House and Senate during the spring of 1874, only to have President Ulysses S. Grant employ a pocket veto that quietly killed it the following year. There seems little doubt that this was a political tactic inspired by the Bureau of Indian Affairs and by General Philip Sheridan.

In fact, Sheridan had clearly laid out a scheme whereby "the extermination of the buffalo herds by meat hunters for the construction gangs of the railroad, by hide-hunters, and finally by tourists shooting from palace-car windows" could hasten the submission of the Native Americans (Slotkin 1992: 426). In a frequently quoted 1875 speech to the Texas Legislature, Sheridan denounced efforts to save the buffalo,

explaining that "the hide hunters were doing more to settle the Indian question than the entire Army had done in thirty years, by destroying the Indians' commissary" (Garretson 1938: 128).

If the American bison were not already doomed to near extinction by 1875, the failure of Fort's bill certainly brought it one step closer. Although within a decade a number of western states enacted laws to halt the slaughter, such regulations were rarely enforced, and even had they been, the laws probably appeared too late to reverse the decline of the species. In 1880, the Northern Pacific Railroad extended its lines through North Dakota, thus offering hide hunters ready access to the still-flourishing northern herds. Within three years, almost all of this population was destroyed.

By the end of the century, free-roaming buffalo had disappeared from the Great Plains and could be found only in a few protected situations: national parks, game reserves, and captivity. By 1913, when the "buffalo nickel" 5-cent coin (designed by James Earle Fraser) entered circulation, wild American bison had vanished from the landscape.

The Struggle for Conservation

The capability of the bison to breed in captivity was a critical factor in its escape from extinction. Following the demise of the wild herds at the end of the nineteenth century, a series of conservation measures were undertaken to preserve the species by controlled management of the few survivors. The initial attempt at government-sponsored protection dates from the beginning of the twentieth century and occurred at Yellowstone National Park, which contained the scant remnants of the northern herd. Although Congress had outlawed all buffalo hunting within the park, poaching was widespread and carried no serious penalty before 1894.

Far greater attention was accorded the buffalo following the 1902 appointment of Charles J. Jones as the warden of Yellowstone. Jones increased the size of the bison population by purchasing privately held animals from throughout the United States. Moreover, he segregated the animals into a "wild herd," which functioned with minimal supervision, and a "tame herd," which was closely monitored and given food and protection from deep snow during the winter. Newborn calves were frequently transferred from the wild to the tame herd to ensure their survival.

The publicity given to Jones's efforts subsequently resulted in the formation, in December 1905, of the American Bison Society (ABS). With Theodore Roosevelt serving as honorary president, and its membership composed of both naturalists and sportsmen, the society proved an effective association for the preservation of the buffalo. Working in conjunction with the New York Zoological Society, the ABS advocated the

formation of additional federal herds to augment those in Yellowstone. Between 1907 and 1918, the societies assisted in the creation of buffalo reserves in Oklahoma, Montana, Nebraska, and North Dakota (U.S. Department of the Interior 1977).

In addition to such conservation agencies, many Native Americans have made efforts to reestablish the bison in North America. As early as 1873, Walking Coyote, of the Pend d'Oreille, noted the animals' decreasing numbers and attempted to raise a herd of his own from four calves captured in Montana. During the 1930s, both the Crow and Sioux established herds on reservation lands. More recently, in 1990, the Inter-Tribal Bison Cooperative (ITBC) was organized to assist bison-restoration projects undertaken by 31 tribes in 13 states. In viewing the bison as a natural part of the plains ecosystem, the ITBC argued that a return of the animal to its former home would improve the lives of the area's human inhabitants (Garretson 1938; Callenbach 1996).

The conservation strategies of the twentieth century have been remarkably successful. No longer endangered, the bison population has prospered to such a degree that the animal has entered commercial markets as a food item. During the 1970s, breeders created the "beefalo," an animal comprising three-eighths bison and five-eighths bovine parentage. Overall, this hybrid offered little beyond novelty; nutritionally, the meat was inferior to that of the buffalo, and the animal appeared susceptible to an array of diseases. The failure of this experiment suggests that, at least in the near future, the American bison is safe from any further genetic tinkering (National Buffalo Association 1990).

Bison as Food

As livestock, the bison provides a high yield of marketable meat, as well as a nutritious final product in the oven or on the stove. A major disadvantage in raising buffalo, however, is the high start-up cost: Prices of bison calves start at around \$1,000, and because buffalo are classified as "exotic animals" by the U.S. Department of Agriculture, aspiring ranchers have less access than cattlemen to government assistance. One considerable advantage is that bison require less food, and far less water, than cattle. Although similar to other bovines in being ruminants capable of digesting other fodder, bison do not depend on succulent grasses. Instead, they prefer short, dry grass, which usually contains a higher percentage of protein. When grazing, bison will consume about 180 pounds less grass per month than cattle of comparable size.

The biggest advantage of the buffalo, however, is its high yield of nutritional meat. With lighter hides, heads, hooves, and tails than Hereford cattle, the buffalo provides a larger percentage of salable meat when dressed (Rorabacher 1970). In addition, when

compared to domestic cattle, it is surprisingly disease free (although respiratory ailments, anthrax, and brucellosis appear as occasional illnesses), and as a result, the meat lacks most of the antibiotics and medications contained in commercial beef and is nonallergenic (National Buffalo Association 1990).

Unlike almost all the meat of other types of domesticated livestock, buffalo meat is low in fat (5 grams per half pound of raw bison, compared to more than 18 grams of fat in the same amount of beef). Clearly, in light of concerns for nutritional health, the bison has great potential as a healthful alternative to most other meats. In areas where butchered bison is readily available (primarily in, but not limited to, the American West), the meat is sold in cuts similar to those of beef: steaks, ribs, roasts, and ground meat. Although the per-pound cost of bison for consumers is often 50 percent greater than, or even double, that of beef, the lower ratio of fat means less shrinkage, and consequently, a pound of bison yields substantially more protein than an equal measure of beef or pork.

The nutritional argument notwithstanding, however, there are an assortment of economic and biological factors that will interact in any determination of whether the bison will again become a significant source of food. Currently, the species is raised on small- to medium-sized ranches in herds of 100 to 250 animals for large-scale production. But the buffalo would doubtless have to be fitted into the beef industry model.

Yet feeding, grazing, meat storage, and refrigeration are only a few of the practices within the cattle industry that would probably require modification to accommodate the buffalo. For example, bison are ill suited to feedlots – facilities used by cattlemen for the rapid increase in the size of their animals immediately preceding slaughter. The confinement created by such ranching tends both to increase fighting and to accelerate the spread of disease among bison. Buffalo, then, cannot be harvested with the same efficiency as cattle. Moreover, although proponents of bison exalt the animals' "natural" meat, free of the vitamins and antibiotics common in other foods, it seems possible, at least, that such additives in buffalo meat would be an eventual result of large-scale commercial production.

It is ironic that economic and cultural change, the same forces that nearly destroyed the bison during the nineteenth century, may hasten the ascendancy of the species in the future. For Native Americans, the buffalo was a sacred resource to be utilized according to human need. However, as the original inhabitants of North America were displaced by white settlers, the hunting of bison became an industry that rapidly diminished the animals' numbers.

At present, North Americans are involved in a cultural shift of sorts: Past understandings of what constitutes good nutrition are being rejected, and many people are reluctant to continue to consume the

high-fat foods that have traditionally composed their diets. Low-fat red meat from the buffalo appears to be a viable replacement for much of the beef presently produced in America. Consequently, once again, harvesting the animals may offer vast financial rewards to humans. However, it is not difficult to envision large-scale bison ranching resulting in a noticeably different meat product from that presently provided by moderate-sized facilities, perhaps even in an unrecognizable buffalo.

J. Allen Barksdale

Bibliography

- Berger, Joel, and Carol Cunningham. 1994. *Bison: Mating and conservation in small populations*. New York.
- Binkerd, E. F., O. E. Kolari, and C. Tracy. 1977. Pemmican. *Maricopa Trails* 1: 1-20.
- Callenbach, Ernest. 1996. *Bring back the buffalo! A sustainable future for America's great plains*. Washington, D.C.
- Catlin, George. 1842. *Letters and notes on the manners, customs, and condition of the North American Indians*. London.
- Cronon, William. 1991. *Nature's metropolis: Chicago and the great west*. New York.
- Dary, David A. 1974. *The buffalo book: The full saga of the American animal*. New York.
- Dobak, William A. 1996. Killing the Canadian buffalo, 1821-1881. *The Western Historical Quarterly* 1: 33-52.
- Dodge, Richard I. [1877] 1959. *The plains of the great west and their inhabitants*. New York.
- Garretson, Martin S. 1938. *The American bison: The story of its extermination as a wild species and its restoration under federal protection*. New York.
- Grinnell, George Bird. 1892. The last of the buffalo. *Scribner's Magazine* 3: 267-86.
- Mayer, Frank H., and Charles B. Roth. 1958. *The buffalo harvest*. Denver, Colo.
- McHugh, Tom. 1972. *The time of the buffalo*. New York.
- National Buffalo Association. 1990. *Buffalo producer's guide to management and marketing*. Chicago.
- Rorabacher, J. Albert. 1970. *The American buffalo in transition: A historical and economic survey of the bison in America*. Saint Cloud, Minn.
- Schultz, James Willard (Apikuni). 1962. *Blackfeet and buffalo: Memories of life among the Indians*. Norman, Okla.
- Slotkin, Richard. 1992. *Gunfighter nation: The myth of the frontier in twentieth-century America*. New York.
- Stefansson, Vilhjalmur. 1956. *The fat of the land*. New York.
- U.S. Department of the Interior. 1977. *The American buffalo*. Washington, D.C.
- Vestal, Stanley. 1952. *Queen of cowtowns: Dodge City*. Lincoln, Neb.
- Wallace, Ernest, and E. Adamson Hoebel. 1952. *The Comanches: Lords of the south plains*. Norman, Okla.
- Wheeler, Homer W. 1923. *Buffalo days; forty years in the old west: The personal narrative of a cattleman, Indian fighter, and army officer*. Indianapolis, Ind.

II.G.2 Aquatic Animals

Pre-Christian Origins of Aquatic Husbandry

Primitive Fish Farming in China

Aquatic farming, like many other modern technologies, is credited to early Chinese societies well before 1000 B.C. In its rudest form, some systematic sowing and harvesting of fish in China have been inferred from marks on ancient “oracle bones” that have survived. Simple shapes of fish were perceived to predict favorable times in which to gather fish “seeds,” particularly those of the common carp that abound in China’s great rivers when the rainy seasons begin, and to sow them in convenient water bodies in floodplains nearer home.

More detailed records of fish husbandry are not found until early classic writings of the Chou Dynasty (1112–221 B.C.). It is not known, however, whether such husbandry was (1) in response to the symbolic cultural and social significance of fish in China, or (2) because of their importance in ornamental ponds built to beautify gardens of the rich, or (3) due to the simple expedient of having fresh food ready at hand each day. In probability, all three reasons are too closely linked to know with any conviction which came first. The common carp has long been a symbol of fortune in China and, therefore, a highly acceptable gift. Indeed, the offering of a live carp (or an image of the fish in jade or ivory rising out of the water to reach the gates of the dragon) recognized the importance of the recipient and bestowed great honor. It is not hard to imagine that both givers and receivers needed a place in which to keep fish alive, and ancient hand-painted scrolls often illustrate scenes from domestic life with tranquil exotic gardens and ornamental fishponds.

The earliest reference in Chinese literature to “aqua-husbandry” as a primitive technology is found in the writings of Fan Li in about 500 B.C. He wrote that culturing carp was one of five ways to make a good living in China, and he described techniques for constructing ponds and for breeding, feeding, and maintaining a healthy fish population. Based on Fan Li’s guidelines, simple carp culture for food production flourished for the next thousand years until, quite fortuitously, an event occurred that revolutionized fish culture in China.

In about A.D. 618, during the Tang Dynasty, an emperor whose family name was Li came to the throne. As the name for common carp in Chinese sounded like “lee,” the idea of culturing, killing, and eating “lee” was sacrilegious and, thus, banned. The crisis, however, compelled farmers to raise other fish and employ new husbandry practices. Soon four new species were cultured successfully, namely the “silver,” “bighead,” “mud,” and “grass” carps. Furthermore,

because of the very different nutritional niches that these fishes occupy in aquatic habitats, the four species could be cultured together in the same pond without competing for food, all of which significantly intensified unit production.

For the husbandry of these new species, farmers soon came to grasp the importance of using organic wastes to increase productivity, and fish culture, as a consequence, became more closely tied to animal and plant husbandry: Manure from pigs and poultry was deposited in fish ponds, and plants and vegetables were grown along their banks. Where ponds could be drained, the enriched sludge was excavated and used as fertilizer for crops. Yields increased dramatically, and polyculture of carp in these integrated farming systems became a key part of rural life in China. Today China annually produces more than 4 million tonnes of freshwater fish almost entirely through the polyculture of eight cyprinids and two or three other species. This figure nearly equals that of all aquatic farm production in the rest of the world.

The spread of fish culture throughout Asia is also credited to the Chinese and to their desire to give tribute, rather than to receive it or to trade. Powerful Chinese naval expeditions moved throughout every inhabited land bordering the China Sea and Indian Ocean for some 500 years before the Great Withdrawal of A.D. 1433, spreading word of the grandeur of each new dynasty, distributing treasures, and teaching skills. Because of their symbolism as valuable gifts, carp, together with simple lessons in fish husbandry, were probably a part of such largesse, especially because the fish are hardy travelers and tolerant of very poor conditions.

The Vivaria of Other Great Cultures

Simple fish husbandry was also practiced in most of the other great cultures. The Egyptians, for example, although not renowned fishermen, built vivaria as adjuncts to their temples and palaces. Their writings indicate that they understood the principles of migration and breeding of fish in the Nile, but there is no evidence of any attempted domestication. Although a staple in the commoners’ diet, fish were eschewed by Egyptian kings and priests for symbolic reasons. Some fish were sacred, as they guided the boats that bore the dead to eternity, and others (probably muddy-tasting freshwater species from the waters of the delta) were regarded as unclean.

The Assyrians, on the other hand, including both Sumerians and Babylonians, were greatly addicted to fish and were skilled fishermen. They were known to keep large resources of fish in lakes created by the construction of irrigation dams on the Tigris and Euphrates, and they built vivaria for their temples and in most townships. Records from 422 B.C. describe fishponds belonging to rich merchants and civic prefects, operated for them by individuals under contract for one-half talent of silver and a daily supply of fish.

Vivaria in Sumerian temples can be traced back to 2500 B.C. Each had a keeper, who charged the public to fish these "piscinae" or stew-ponds. Further evidence indicates that piscinae became quite numerous among Assyrian commoners, although it is recorded that ponds built by "poor men" often drew poachers because their owners had no legal redress.

The neighboring Israelites, in spite of their liking for fish and their links with Egyptians and Assyrians, apparently never constructed vivaria. There is no record of such activity in all of their prolific writings until just before Christianity was born, when the Israelites were probably influenced by the practices of the Romans.

Wealthy Greeks and Romans feasted on all the common marine fish and shellfish of the Mediterranean for which, at times, they paid extraordinarily high prices. The Greeks appear to have been satisfied with what they could catch conveniently close at hand, and there is no record of their building vivaria. The Romans, on the other hand, transported seafoods, such as mullet, oysters, and cockles, great distances inland at significant cost and brought back species to Italy from all parts of their empire.

As a consequence, the Romans adopted the Egyptian and Assyrian vivarium, which became a necessary component of all the great estates. Many were constructed indoors, often adjacent to refectories to enable dinner guests to see and choose their own fish, and some were even built on board ships. Although costly to construct and maintain, vivaria proved extremely popular, and what had been intended to keep food fresh in a hot climate became a prestigious showpiece as well, complete with keeper.

The Roman vivaria remained the prerogative of the wealthy. However, the rich brackishwater lagoons and "valli" that surrounded the coast of Italy served the common people in similar fashion. They were especially good sources for mullet, sea bass, sea bream, and eels, which could be trapped inside them.

Whether the practice of keeping fish in vivaria was carried by the Romans throughout their empire is not known. The foundations of excavated Roman houses and fortifications outside Italy reveal pits that held water, but these may have been only for storing drinking water. Probably the defensive moats of larger walled fortifications and cities contained water enriched with human and kitchen wastes, which fell from the walls above and provided good nourishment for certain types of freshwater fish, such as carp. But, on the whole, freshwater fish were not popular with the Romans.

Subsistence Practices of the Middle Ages

Fish Ponds of Tribal Societies

Outside of China, early aquatic farming for food production developed mostly through tribal societies who fished for subsistence and survival. In most such

societies, fishermen relied on a variety of fixed traps, in addition to spears and lines with hooks fashioned from bones for traditional fishing. These traps ranged from simple woven and baited baskets suspended in the water, to earthen ponds constructed along the shoreline, to complex labyrinths of fences made from bamboo and covered with reeds that enticed fish into ever-receding spaces from which there was no escape. Doubtless such fishermen also discovered that most fish could be held in captivity in these traps and kept fresh for several days or even longer if some feed was provided. From this discovery it was only a short step to locating places where seasonal resources of fry and fingerlings could be captured in large numbers, impounded in some way, and raised for future harvest.

As exemplified by surviving primitive societies, tribal fishermen have a good understanding of the natural behavior of their quarry and the influence of seasonal and lunar changes and diurnal conditions on their movements. Thus, the trapping and holding of brackishwater fish, such as milkfish and mullet, believed to have started in Indonesia in the fifteenth century, was possible because it was understood that every year, millions of young fish would migrate inshore in early spring to mangroves rich in food. Here they were caught and transferred to shallow ponds crudely fashioned around the estuaries and kept filled with water.

Records show that brackishwater ponds were also an important part of Polynesian societies throughout the Pacific islands at about the same time and that pond practices were quite sophisticated. The earliest recorded date for the construction of fishponds in the Hawaiian Islands is in the middle of the fifteenth century. These tidal ponds, shaped by stone and coral walls on top of reefs, were built around or adjacent to streams so that the fresh water attracted and fed young fish.

Most of the ponds had a system of one-way gates, through which water could be exchanged and small fish moved in (to feed and grow), and by which large fish were trapped at the full moon (as they tried to move out to spawn at sea). Other ponds had no gates, but water was exchanged through the permeable coral walls.

Ancient Polynesian walled fishponds were owned by chiefs and were symbols of importance, built by people drafted for the purpose. By the beginning of the nineteenth century, more than 200 ponds were recorded around the islands of Hawaii. Many were large, with walls from 2,000 to 5,000 feet in length. Each pond had keepers who fed the fish with taro and other vegetation and caught them for the chiefs and their immediate retinue.

Because the common people had no rights to fish in these ponds, they made their own "underwater ponds" in the sea by connecting coral heads with walls, thus trapping fish as the tide receded. They

were also allowed to operate fishponds inland in wet areas where taro was grown. The Polynesians were a marine-dependent society, and the large variety of fish abundant around their islands in the Pacific was the principal animal protein in the everyday diet. They were not dependent on fish production in their ponds for subsistence, but used them as a source of food in bad weather and as a ready supply for feasts.

Coastal fishponds were probably built in Asia well before the fifteenth century. The islands of the Hawaiian chain and Tahiti, which lie almost at the limit of Polynesian extension in the Pacific, are known to have been well populated by A.D. 1000 by means of migrations of people descended from the Melanesians of the far western Pacific. Therefore, it is probable that the skills of catching fish and keeping them in captivity were carried from China, through Southeast Asia, into the Melanesian islands, and then on out into the Pacific between the tenth and thirteenth centuries.

The Stew-Ponds of Europe

At the same time that it was extending throughout Asia and the Pacific, the knowledge of fish husbandry also traveled westward into Europe. In the Middle Ages, so-called stew-ponds became increasingly common on large estates. They were a convenient source of fresh protein to alternate with a diet of dried, salted, and pickled foods, and particularly important in times of war and siege.

In similar fashion, coastal beds of common shellfish, such as oysters and mussels, were maintained as reserves of fresh protein. These were protected and conserved by selective harvesting, but there is no evidence of any effort at true domestication.

As with earlier cultures, stew-ponds in western and eastern Europe were associated with priests and nobles. They are evident in the earliest records of many of the religious orders founded in the twelfth century. For example, the original charter of the Kladruby Monastery in Bohemia, dated 1115, describes a fish production pond, and many others are still visible in the ruins of the earliest monasteries.

Stew-ponds were also a common feature of the great estates – most of which were constructed beside fish-laden rivers. In England, the eleventh-century Domesday Book records ownership of “vivariae piscinae” by many large estates and also by smaller landholdings of ordinary country squires and wealthy middle-class merchants. In his prologue to the *Canterbury Tales*, written in 1387–92, Geoffrey Chaucer describes the epicurean Franklin, a “householder” who “hadde many a breem [bream] and many a luce [pike] in stuwe.”

In contrast to such largesse, the feudal system of the European Middle Ages denied landownership and, therefore, stew-ponds, to peasants. All rivers and streams belonged to kings and their barons, who controlled large territories, as well as the game and fish in

them. Poaching in rivers as well as in stew-ponds was generally punished by death, and thus fresh fish was a rare commodity for the lower classes and rural peasants. When ancient laws regarding fish and fishing were finally changed (some such statutes remained until the end of the nineteenth century, as in Hungary), the establishment of family stew-ponds was one of the first products of such change.

Toward the end of the Middle Ages, the fish typical of stew-ponds in Europe were native freshwater species, such as bream, perch, carp, barbel, roach, dace, and minnows. Predatory fish, such as pike, tench, eel, and lamprey, were excluded if possible. With the exception of such migratory fish as salmon, those which remained through the winter would probably reproduce in the following spring if the pond was suitable and large enough – thus replenishing the stocks naturally. From this point it was but one more step to the discovery that separate ponds were useful for storing during winter, breeding in spring, and fattening in summer and autumn.

By the middle of the fifteenth century, the construction and management of fishponds was a well-established and integral (but small) part of artisanal life, particularly in eastern Europe. Many towns in Bohemia and Moravia competed with the monasteries and nobles in pond construction, some of which were extraordinarily large. About 20,000 ponds were registered in the archives of this region in the sixteenth century, covering more than 75,000 hectares.

The technology, particularly of carp culture, became quite advanced. Many ponds were managed intensively, with production increasing steadily to between 75 and 100 kilograms per hectare. Techniques were described in books of the period. In 1547 Janus Dubravius, the Bishop of Olomouc, produced his treatise, *Jani Dubravil de Piscinis et Piscium qui in Illis Aluntur Libri Quinquae*, later translated into many languages. It included chapters on economics and diseases and described the work of fish-culture wardens and their organizational guild. Other works by John Taverner (1600) and Gervais Markham (1613) discuss in detail the practices of fish raising, including pond construction, pond fertility, management, feeding, and the best fish to raise. These authors also made observations on the breeding behavior of adults and the care of fry and fingerlings.

The beginning of the Renaissance coincided with the first peak in the early domestication of aquatic animals worldwide. By this time, thanks to the skills of the Chinese, most countries of Asia practiced breeding and propagation of several freshwater fish in captivity and had developed relatively sophisticated production practices. Island communities of the South China Sea and the Pacific Ocean substituted marine fish that could not be bred but that could be gathered in large numbers to stock coastal ponds.

The Europeans had also steadily advanced in management practices for the husbanding of freshwater fish in captivity and relied on natural breeding to supplement their stocks. But the next two hundred years saw stagnation and even decline in aquatic animal domestication. China closed its doors on the world, whereas in Europe there was growing competition for water and for land (Radcliffe 1926; Toussaint-Samat 1992).

The Impact of the Industrial Revolution

The Demise of Inland Fisheries in Europe

At the dawn of the nineteenth century in Europe, the demand for fish and shellfish was met by the traditional and dominant inland fisheries and by some modest estuarine and coastal fisheries. This adequately supplied a population widely dispersed among rural villages or grouped in small agricultural towns and a few principal cities that were trading centers.

Nineteenth-century industrialization, however, which mechanized fishing fleets and opened up large urban markets for iced fresh fish through networks of railways, expanded modest coastal fishing into marine fishery industries at unprecedented speed. Unfortunately, the same industrialization decimated the traditional freshwater fisheries.

The freshwater fisheries of coastal Europe were based mostly on migratory species, predominantly Atlantic salmon, sea trout, and, to a lesser extent, the eel. However, the Industrial Revolution produced an insatiable demand for fresh water, and almost every river and stream was affected. They were dammed to make reservoirs and drained by growing towns, polluted by manufacturing industries built along their banks, and diverted to fill new canals, and their catchments were stripped bare by deforestation.

Furthermore, those fish running the migratory gauntlet were poached with every conceivable trapping device, including explosives and poisons. Although protective fishery laws had been in existence since the fourteenth century, they were largely ignored and not enforced. Finally, adding to this relentless assault on the migratory fisheries, many large lakes and ponds of central Europe – operated since the Middle Ages as inland fisheries for the common carp and other cyprinids – were drained to meet the need for more agricultural land.

In a mere hundred years, and before parliamentarians finally recognized the problem and took action, the Industrial Revolution had polluted all the large and most productive estuaries and rivers in Europe or made them otherwise impassable to fish. Most remain so today. It is fortunate that toward the end of the nineteenth century, public and private organizations were formed to safeguard and monitor inland fisheries and to undertake scientific research to bring about their recovery.

Artificial propagation to replenish falling stocks was one small part of their regulatory response for the management of all fisheries, but this was not the only reason for such development. Another was a growing scientific curiosity in natural history inspired by the voyages of Charles Darwin and Alfred Wallace and also the new field of genetics discovered by the monk, Gregor Mendel.

Fish Propagation and the Hatchery Movement

The breeding of common brown trout was first described by William Yarrell in his *History of the British Fishes*, published in 1841. The book was based on a translation of an original paper written in 1763 by a German naturalist, which had gone largely ignored. The technique was rediscovered in 1844 by two French fishermen, Anton Géhin and Joseph Remy, who not only told of successful fertilization of trout eggs but also of stocking the progeny in the Moselotte River and raising them in a pond. Subsequently, Gottlieb Boccius wrote *A Treatise on the Production and Management of Fish in Freshwater by Artificial Spawning, Breeding, and Rearing*, with the sinister subtitle, *Shewing also the Cause of the Depletion of All Rivers and Streams* (1848).

Other works rapidly followed, notably those of Professor M. Coste in France. Following the publication of his *Instructions Pratiques sur la Pisciculture*, he was invited to lecture on fishery problems throughout Europe. These travels culminated in his publication (1861) of the *Voyage d'Exploration sur le Littoral de la France et de l'Italie*, which described the artificial culture of both fish and oysters. In addition to being generally credited with saving the oyster fisheries of France, he was also responsible for founding the first fish hatchery, built at Huningue, near Basle, in 1852. From this hatchery, eggs were distributed throughout France and other European countries to many rivers before these countries built their own hatcheries.

Coincidentally, there were similar developments in the New World. In Ohio, Theodatus Garlick artificially bred brook trout from eggs that he brought back from Canada in 1853. Within the next 20 years, eggs of a number of common freshwater fish had been reared under primitive laboratory conditions and the young fry put into their natural waters before they would have died in captivity.

The majority of these species were common salmonids, such as brook trout, brown trout, rainbow trout, Arctic char, and members of the large group of Pacific salmon, as well as other popular non-salmonids, like the shad. All these fish had common attributes. They were readily caught, their eggs were large and visible, and the emergent fry had a plentiful supply of food from the yolk, which avoided the problem of feeding before release.

Many of these early efforts were supported by landowners and gentlemen sportfishermen. With independent financial resources to build hatcheries, these dilettantes started many small programs. In 1868 the first hatchery for Atlantic salmon was built by Samuel Wilmot on his own property adjacent to Lake Ontario, Canada. There he released parr into the creek, and in 1867, with money from the government, the Wilmot's Creek hatchery was expanded to produce, in time, about 1 million salmon each year. This was, in all probability, the first actual ranching of salmon raised and released from a hatchery.

Soon fish hatcheries were common throughout North America. The first Pacific salmon hatcheries were built on the McCloud River in northern California in 1872 and in the Columbia Basin on the Clackamas River in Oregon in 1877. By the end of the century, the state of Washington alone had 14 salmon hatcheries, producing more than 58 million fry.

Experiments seemed to know no bounds. Rainbow and brook trout were transplanted from the western United States to trout hatcheries in Europe; and in the mid-1860s fast clippers carried the first shipments of trout and salmon eggs to India, New Zealand, and the new Chitose hatchery in Japan. In the early 1870s, Denmark built a hatchery to raise rainbow trout recently introduced from the United States. However, instead of releasing the young fish to enhance local rivers as others were doing, the Danes continued to keep the fish in captivity. This became the first land-based salmonid farming enterprise and led to Danish domination of the trout industry for the next hundred years.

The great 1883 International Fishery Exhibition in London, which devoted considerable space and attention to the new technology of fish (and shellfish) culture for enhancing declining fisheries, proved to be the springboard for worldwide interest. Many countries, including China, provided wonderfully illustrative and mechanical exhibits, and scientists and public figures alike returned home resolved to implement newly gained ideas.

One such man was Lachlan Maclean, whose successful introduction and acclimatization of trout in South Africa (although achieved only by persistence and endurance) led to other such attempts in all European colonies and overseas territories wherever there were cool mountain streams. Thus, trout were stocked in suitable upland areas throughout East Africa, from South Africa north to Kenya, and transferred to Australia, Tasmania, and New Zealand. They were also shipped to the Andean countries of South America.

In the end, however, most of these essentially private enterprises were not successful. There was still a great lack of understanding about the behavior, reproductive biology, nutrition, and diseases of salmonids; hatcheries were not simple facilities to construct and operate, and fish culture was not cheap. But, fortu-

nately, by this time many governments had established fisheries policies, and fish propagation was an active part of fisheries management. Consequently, by the end of the century almost all private hatcheries had become publicly operated.

The First Propagation of Marine Species

In 1867, following an extensive survey of the oyster fisheries of Europe, the British naturalist Frank Buckland broadened his interest in freshwater perch to include oyster culture. His goal was not directly to produce food, but rather to use artificial culture of aquatic animals to enhance natural fisheries.

Several influential scientists, however, including Thomas Huxley, stated that farming the sea was useless as the sea fisheries were inexhaustible. But, as noted by one of Buckland's proponents, Ernest Holt, this was an overgeneralization. He believed that although the great fisheries for cod and herring and pilchard were beyond the influence of humans (either by overfishing or restocking), there were several smaller but important fisheries in the North Sea, such as that for flatfish, which could be enhanced. Holt showed with statistics the effects of overfishing on the plaice fishery, and in 1897 he proposed four solutions for management - one of which was artificial propagation.

The activity in fish propagation and fisheries research, triggered by the Great Exhibition of 1883, made it clear that temporary facilities for marine culture in fish-processing plants and water mills were totally inadequate. Propagation needed hatcheries, and fishery science needed institutes and laboratories. Consequently, one final contribution of the century to fish culture was the founding of many of today's most famous marine research stations, along with several professional societies.

France was one of the first countries to construct facilities directly in support of marine fish and shellfish culture. Following the successful efforts of Coste in salvaging the French oyster fisheries, coastal centers were built at Concarneau in 1859 and at Arcachon in 1863. In addition, a laboratory was built at Monaco, and in Naples, Italy, the famous Stazione Zoologica was constructed.

In the United States, Spencer Baird was appointed the first Commissioner of Fish and Fisheries in 1871, and fish culture became an official responsibility of the government. The first facility for work on the Atlantic cod was built in Gloucester, Massachusetts, and propagation was successfully achieved in 1878 using the technique developed by the Norwegian G. O. Sars in 1866. The successful propagation and release of other gadoids, and even herring, quickly followed, despite what would be described today as relatively poor conditions. Government funds were provided in 1885 to construct the first commercial marine fish hatchery at Woods Hole, where there was

also a small group of scientists carrying out research on marine fish. The results were so good that the government built a second hatchery at Gloucester Harbor.

In the United Kingdom, the Marine Biological Association of England was founded in 1884 by the Royal Society of London, under the presidency of Thomas Huxley. In its first resolution, the society emphasized the necessity of establishing one or more laboratories on the coast of Britain where research could be conducted that would lead to the improvement of zoological and botanical science, and to an increase in knowledge about the resources of the sea.

The first laboratory was opened at Plymouth in 1888 with three professional staff, followed by one in Lowestoft. The new Lancashire and Western Fisheries Committee built laboratories at Port Erin on the Isle of Man and another at Piel Island near Barrow-in-Furness, both under the direction of Professor William Herdman from Liverpool. He began substantial propagation and release work not only with the flatfishes (plaice and flounder) but also with Atlantic cod and haddock.

The first marine hatchery in Scotland was built in 1894 at Dunbar, near Edinburgh, under the direction of James Cossor Ewart. The purpose of the Dunbar hatchery was to raise and release young marine fish to add directly to the fish supply. It began with the propagation of plaice to enhance the coastal fisheries. Ewart was helped by Harald Dannevig who, with his father Gunnar, had built the first commercial hatchery for Atlantic cod in Norway in 1882. Other facilities were built at St. Andrews and Aberdeen, and the newly formed Scottish Marine Biological Association constructed its new laboratory at Millport in the Firth of Clyde.

The fishery science carried out at all these coastal laboratories was well structured, largely because in 1899 and 1901, a gathering of illustrious scientists at Christiana, Sweden, established common methodologies for gathering information and identified specific cooperative programs for biological and hydrographic studies. The "Christiana Programme," as it became known, coordinated some of the first fisheries and marine research over an extensive area of the North Sea, the Northern Atlantic Ocean, and the Baltic Sea.

Early marine laboratories and fish hatcheries had large outdoor ponds or tidal basins for holding broodstock, modeled after those designed by the Dannevigs for cod. But it was soon clear that better control was required for incubation and for subsequent larval rearing, and in the United States initially, the "Chester jar" and then the larger "MacDonald tidal egg-hatching box" were adopted.

In Europe, the "rocking incubator" developed in Norway was more popular. But common to all hatcheries were cumbersome mechanical water systems, all soon blocked by biofouling. Worse still, their metal pipes and valves rapidly corroded, filling the tanks

with rust and dangerous toxic ions. New hatcheries were built with dual pipe systems, settling tanks, and sand filters, substantially increasing their costs. Nevertheless, many such problems were not resolved for almost another half century with the advent of inert plastic materials.

Because they were concerned only with the early life stages of aquatic animals, marine hatcheries all over the world had no difficulty in realizing some large production numbers. Three hatcheries in Massachusetts, for example, accounted annually for 3,000 million fry of pollack and flounder. In Australia, under the expert eye of Harald Dannevig (who emigrated there in 1902), the hatchery at Gunnamatta Bay in New South Wales was producing 150 million fry annually; and this was emulated by the new Dunedin hatchery in New Zealand.

Unfortunately, however, despite these massive numbers, the releases had little beneficial effect on the local fisheries. This was because, unlike the eggs of freshwater fish, those of marine species are small and have little in the way of yolk reserves. The eggs hatch out quickly and the emergent larvae must feed almost immediately or die. Culturists knew little of the first feeding requirements of the emergent larvae and consequently, as there was no effort to provide the fry with small live-food organisms, their survival was almost immediately jeopardized.

The First Half of the Twentieth Century

The Influence of the Colonial Empires

In spite of the rapid progress in aquatic husbandry during the final 20 years of the nineteenth century, the first 50 years of the twentieth century were spent marking time. This was a period when efforts, intentional or not, went into building infrastructure, rather than converting scientific and technical advances into commercial production. But it was also a period when colonial and territorial influences in Asia and Africa were at their peak, and these had several useful benefits for the spread of existing technology around the world.

The new field of fisheries, and especially the idea of introducing and stocking familiar sports fish, was of considerable interest to many colonial administrators, who established fisheries departments, appointed fisheries officers, and built research centers and government farms. In many cases, an administrative organization was repeated at the state level. In India, for example, the first government fish farm was established in 1911 in the state of Sunkesula by the Fisheries Department of Madras, and this was followed by stations in Bengal, Punjab, Uttar Pradesh, Baroda, Mysore, and Hyderabad.

In India, not all of this activity was directed at the culture of sports fish, however. Eating meat was forbidden to large parts of the population by religious law, and meat was a scarce and costly commodity for

many others. Hence, the possibility of increasing traditional fish production through culture was an attractive option. The gourami (first brought to India in 1841), Chinese carps, and some tilapia species were either introduced or reintroduced and became the basis of considerable practical research.

With its cultural background and strong tradition of fish and shellfish consumption, Japan was one of the earliest countries to establish regional research stations in support of national fisheries. The first was at Aichi, built in 1897, followed in the next 12 years by Shizuoka, Fukuoka, and Kuwana. This program was assisted by one of the earliest pieces of legislation in support of aquaculture, namely the Reclamation Subsidy Act, which designated large areas of land for reclamation for fish culture ponds. New ownership of these lands encouraged the capital investment required to develop them for eel culture, which increased the demand for wild elvers. The situation was further enlivened by a sudden recession in agriculture, which galvanized many farmers to convert their lands to fishponds.

The Japanese occupation of part of China, including Formosa (Taiwan), which was ceded by China in 1895, was accompanied by the transfer of Japanese interest in aquatic culture to the island. The Governor-General of Formosa established and organized an infrastructure for fisheries, much as it existed in Japan, and constructed a fish culture station at Tainan in 1910 under the distinguished scientist Takeo Aoki. Building first on the culture of milkfish (introduced centuries before by Dutch occupiers), the Japanese scientists worked on a variety of species. These included key marine species, such as the yellowtail, sea bream, and abalone, as well as freshwater eels and tilapia introduced from Indonesia. They built several other stations throughout the island, all of which were retroceded to the Republic of China in 1945.

In the colonial countries of Africa, particularly those belonging to Belgium and Great Britain, the priorities were for producing sports fish for recreational fishing and mosquito-eating fish for controlling malaria. The first recorded introductions of sports fish for culture in Africa (in particular, salmonids for cool mountain streams) were mostly confined to the rich agricultural countries with suitable upland areas. These stretched from Kenya down to South Africa and included Uganda, Nyasaland (Malawi), Rhodesia (Zimbabwe), and Swaziland. However, records also reveal the introduction of many other species, such as Chinese carps, and the liberal transfer of many native species, such as the cichlids.

In the north, as part of its continuing irrigation program in the Nile Delta, Egypt built the Barrage Farm in 1919 to receive introductions and produce seed for stocking its coastal lakes. The El Mex Farm was constructed in 1931 at the pumping station near

Alexandria, and a number of marine species, such as Dover sole and shrimps, were introduced into inland high-saline waters, such as Lake Qarun.

Conservation and Compensation in the New World

Because of a lack of tangible benefits, North American interest (and investment) in the hatchery propagation of marine species was waning rapidly by the 1920s. But in the United States it was on the increase for freshwater species – especially Pacific salmonids – and, fortunately, the techniques of propagation were now highly reliable. It was fortunate because of the crisis that was beginning to develop in the salmon fisheries of the Pacific Northwest, and nowhere was the crisis more keenly felt than in the vast area of the Columbia River watershed.

The fisheries of the giant Columbia River had supported some 50,000 persons and yielded about 18 million pounds of fish each year. However, a number of factors combined to threaten the salmon resources of the Basin. Among these were the opening of the Pacific Northwest territories at the end of the century, the construction of hydroelectric dams, the diversion of water for irrigation, and poor logging and mining practices.

Fish hatcheries and fish ladders were constructed in the region to mitigate some of the losses, but it was not until the Federal Power Act of 1920 that fishways were required at all private power projects. The Fish and Wildlife Conservation Act of 1934 that followed was the most important legal authority for ensuring both protection and compensation for the salmon fisheries affected by federal water projects. Together, these two acts were responsible for more than 400 million U.S. dollars spent on fish passages and propagation hatcheries constructed in the Columbia Basin.

World War II

During the war, as the North Sea was closed, near-shore fishing was concentrated along the safe west coast in the British Isles. It was proposed that fish production could be supplemented by increasing productivity of some Scottish sea lochs with inorganic fertilizers and by increasing resources of fish through the transplantation of juveniles, particularly flatfish.

Led by Fabius Gross from the University of Edinburgh, a team of scientists began the fertilization of Loch Craiglin with sodium nitrate and commercial superphosphate fertilizers supplied by Imperial Chemical Industries Ltd., the largest national producer. The fertilizers increased productivity in the loch, but most of the nutrients were taken up by the phytoplankton and algae. Food organisms suitable for growing flatfish were also increased, but not with the consistency required to support large populations.

The economics of one of the first true attempts to farm the sea were never tested, however. As fishing

vessels began to return to the seas in 1944 and 1945, the group was disbanded, and with the early and untimely death of Gross in 1947, all further technical interest was lost.

Another event concerning aquatic husbandry during the war years occurred in Africa, which was a major resource of raw materials for the allies. Mining in South Africa, and in other countries such as Zambia and Zaire, was greatly expanded, and large pools of labor worked the mines, built railways, and manned ports to maintain the flow of copper and iron ores. This labor force had to be fed and, for a time, meat was secured by large hunting parties and by deliveries of beef from South Africa.

It was not long, however, before local herds of wild game had been depleted and, to make the situation more critical, the beef was being redirected toward armies and to the hungry of Europe. British and Belgian colonial administrators reacted to the crisis by ordering the construction of ponds to produce fish for the miners and their families. Many thousands of small fish ponds were dug and stocked with tilapia and other native species for food. The effort proved successful and, at the end of the war, the Conference Piscicole Anglo-Belge was organized for further post-war development of aquaculture for food production.

A third event involving aquaculture had even more far-reaching importance. Fish farming in what would become Israel had been developed in 1934 by immigrants from central Europe, who brought with them experience in the culture of their traditional (and prized) fish, the common carp. The first kibbutz fish farm was established in 1938 at Nir David, and by the end of the war there were 30 productive kibbutz farms with a water surface of 800 hectares.

Initially, traditional methods were used on the farms, but as land and water became scarcer, new systems and practices were developed that were unique and highly productive for arid lands. In time, these would make Israel self-sufficient in fish production and even an exporter. In addition, the program made the country one of the most progressive in the field of fish culture.

The Birth of Modern Aquaculture

The Postwar Years

The wartime interest in fish farming for food production dimmed with the end of the war. Without extensive harvesting for many years, coastal and oceanic fisheries were ripe for such an effort, and world harvests increased steadily as old fleets of traditional fishery nations were modernized and expanded. In 1950, a world harvest equaling that of 1938 (about 20.5 million tonnes) was realized and this was soon doubled.

The early postwar years, however, were years in which many developing countries were newly independent with swelling populations, and "food self-suf-

ficiency" and "animal protein" became prominent problems. European governments continued to contribute significantly to increasing food resources through production schemes and direct technical assistance. This had a valuable impact on the further development of fish and shellfish farming. Science was taken out of the laboratory and applied in the field.

The pioneers in this effort were trained in basic biological sciences and devoted the greater part of their lives to working overseas on almost every aspect of improving aquatic animal production. C. Fred Hickling, for example, like many of his colleagues from the British Museum, first worked on fisheries in East Africa. He studied the genetics of tilapias and produced the first all-male populations. Later he continued his work at the fish culture station at Malacca in Malaysia, which he expanded with its Latin square of research ponds, and where he contributed to advances in soil and water chemistry.

Similarly, Antoon De Bont and Marcel Huet, from Belgium, spent much of their lives in Central Africa and Indonesia, working on the breeding of tilapias, carps, and general limnology, while Jacques Bard, of France, worked in West Africa and later South America on tilapias and many indigenous species. Other such individuals included W. Schaperclaus from Germany and Elek Woynarovich from Hungary.

In the 1950s, many of these pioneers helped the fish culture staff of the Food and Agriculture Organization of the United Nations (FAO) to spread tilapia species throughout Asia and the Pacific region in an attempt to provide easily produced and readily available protein to the rural poor. Unfortunately, the project was to have long-lasting repercussions, as the fish rapidly replaced indigenous fauna and became pests wherever they were introduced.

After World War II, the pressing priority for food in Japan meant that all forms of agriculture received high priority during the reconstruction, and the old technical infrastructure was rapidly replaced, with emphasis on practical production.

Similarly in Taiwan, following independence in 1949, the old Japanese research stations were used to develop fish and shellfish culture under the Joint Commission on Rural Reconstruction, funded by the Rockefeller Foundation. Under the direction of T. P. Chen, fish culture in Taiwan became an important economic industry, principally for the efficient farming of milkfish and the fattening of eels for the Japanese market. There were also successful research developments, particularly the control of breeding and propagation of silver carp, grass carp, and grey mullet and, for the first time, species of marine shrimp.

In the United Kingdom throughout the late 1950s, the government supported two research projects as part of its general program on marine biological research. One, under James Shelbourne at Lowestoft,

was concerned with the breeding and propagation of marine flatfish, whereas the other, under Duncan Waugh at Conway, focused on the propagation of oysters.

In 1961 the British White Fish Authority, a quasi-government organization operated by the fishing industry, instigated projects to carry promising research results from these projects through to commercial implementation. The authority constructed two experimental hatcheries. One, for marine flatfish, was located at Port Erin on the Isle of Man, while the other, for oysters, was situated at Conway. The successful operation of and yields from these hatcheries enabled follow-up efforts to be made at field stations in an attempt to accelerate growth rates. One was constructed in Scotland for the production of flatfish in an enclosed loch at Ardtoe. Another was located at Hunterston in Ayrshire, in the heated effluent from a nuclear electrical generating station.

Shelbourne's successful propagation of marine flatfish in England was possible because of an important earlier discovery made by Gustav Rollefson in 1939. Rollefson had found that the nauplius of the brine shrimp was a useful live-food and small enough for feeding marine fish larvae. Moreover, the eggs of brine shrimp were at times encysted, and they could be stored conveniently for subsequent use.

After the war, the natural production of cysts was first exploited at the Great Salt Lake in Utah for growing tropical fish for the aquarium trade. This was rapidly followed by more controlled production in the solar saltworks of San Francisco Bay. The availability of large and reliable quantities of instant live-food opened the way for marine fish and shellfish production on a hatchery scale, but this time with the potential to produce viable fry instead of fragile, starving larvae.

Uncontrolled Expansion of the 1970s

Throughout the 1960s, work by the White Fish Authority on marine flatfish in England, the growing production of yellowtail in Japan, and the successful farming of eels in Taiwan captured the interest of governments and scientists alike all over the world. The concept of farming the sea, an idea that had lain essentially dormant for more than 50 years, was suddenly reborn, and it was also an idea that reinjected new life into the passive efforts of farming freshwater species.

The 1970s witnessed an explosion of effort by marine biologists the world over. In Norway and England, attention was focused on the production of Atlantic salmon and rainbow trout. In the United States, scientists reinvigorated their work on Pacific salmon and the farming of catfish. Israel, in its continuing search for food self-sufficiency, stepped up its concentration on carp and tilapia production through integrated farming. Hungary produced and bred individual strains of carp and adopted Chinese methods for integrated farming with ducks.

In Japan and Taiwan, work intensified on the production of marine shrimps, followed by efforts in the United States, Panama, Ecuador, Thailand, the Philippines, Indonesia, and England. The Philippines and Indonesia improved their milkfish production. Thailand and the United States pioneered research and development of freshwater prawns. Japan expanded its efforts on abalone, sea bream, mackerels, and edible marine algae. Taiwan improved production of milkfish and bred mullet for the first time. Spain and the Netherlands produced prodigious quantities of mollusks, while France worked on sea bass, flatfish, and oysters, and Italy on sea bass and sea bream. In Europe, Denmark, England, and France revitalized their flagging trout industries, and Germany its eel industry.

Such a decade of intensive effort was assisted in no small way by complementary technologies. Plastics and fiberglass revolutionized hatcheries. Traditional cast-iron and glass piping and large concrete tanks, characteristic of the late nineteenth- and early twentieth-century hatcheries, were replaced by light and easily assembled plastic and fiberglass materials, which were inert and noncorrosive. Floating cages and rafts proved effective substitutes for costly excavated ponds and replaced the need to purchase expensive flat land. And, finally, modern scientific equipment enabled water conditions to be instantly monitored, bypassing lengthy laboratory analyses.

The domestication of aquatic animals was again viewed as a new and promising field of modern agriculture. Governments poured in development funds on promises of a technology to provide cheap protein for rural populations and profits for investors. Professional societies were formed, and the new field was now called "aquaculture," which, unfortunately, would experience still another dip in its progress curve.

The Realities of the 1980s

The optimistic view of the potential of aquaculture generated in the 1970s received a serious setback in the 1980s because when technical achievements in the laboratories were expanded to commercial endeavors, they were not profitable.

The logistics of manufacturing and delivering increasingly large quantities of artificial feed had not been adequately estimated, and such feed often proved to be nutritionally deficient. Thus, large losses of stock were encountered through poor nutrition, which exposed the animals to a variety of new diseases for which there was no treatment.

In addition, there were shortages of skilled labor, and many farms were victims of unforeseen disasters, running the gamut from storms and floods to toxic algal blooms, water pollution, and a general failure of equipment. Consequently, the sporadic quantity and quality of many farmed products were not competitive with those of natural products, and many investments were lost.

The 1980s, therefore, became a period of intensive research. In most cases it amounted to backtracking on matters of nutrition, pathology, and engineering. But research also advanced in new areas, such as induced breeding to improve egg quality and genetics. Work on many species was stopped, and efforts were focused on those species that could be produced reliably and in the quantities to make investments profitable.

Fish and Shellfish as Food

Large resources of protein in the flesh of fish and shellfish contain many readily available amino acids, such as lysine, methionine, and tryptophan, in quantities comparable with those in eggs, meat, and milk. Indeed, with their unsaturated fats, vitamins, minerals, and trace elements, fish and shellfish constitute a near-perfect food for the human body.

Obviously, however, quantitative differences in nutritional composition occur among the many large groups of species, and also within groups at different times of the year. In general, freshwater and brackishwater species contain about 14 to 25 percent protein, whereas marine species contain 9 to 26 percent.

Freshwater species have a low percentage of fat (less than 1 percent), whereas in some marine fish it may be as high as 20 percent. Compared with animal fat, fish oils contain more polyunsaturated fats and, therefore, can be beneficial in reducing the buildup of cholesterol in blood. In addition, fish and shellfish are good sources of calcium and phosphorus, especially when small fish and crustaceans are consumed whole. They also contain iron and traces of copper, as well as B-complex vitamins.

Preservation of Fish

Fish and shellfish offered early cultures a rich source of animal protein. But both were (and are) extremely perishable food commodities, subject to the activity of microorganisms (such as bacteria, molds, and yeasts) and internal chemical deterioration (such as rancidity from breakdown of oils and fats and enzymatic actions), as well as targets of insects and scavengers in storage. Preventive measures to slow down these many processes of deterioration were, therefore, required.

The traditional preservation methods of indigenous cultures around the world today are probably indicative of much trial and error that went on in primitive societies as they perfected their processes and built experience in storing food. Although drying, smoking, salting, boiling, and fermenting have been the five basic traditional methods of preservation or processing, there are many, often subtle, differences in method among and even within cultures. These are due to different preferences for taste, tex-

ture, smell, color of the flesh, and social custom. They are also due to the environment of the consumers, the habitation in which they live and store food, and the availability of materials for processing, such as the fish and shellfish themselves, sources of fuel for the fire, and the availability and composition of salt.

Early fishermen must also have learned the hard way that some species, or their body organs, were highly toxic and often lethal, and that such dangers could occur at different times of the year. For example, several mollusks can cause paralytic shellfish poisoning if harvested and consumed during incidences of intensive algal blooms in hot weather, and some individuals within common reef-fish species of the Pacific Ocean may induce ciguatera poisoning. Parts of the fish can also be highly toxic. The puffer fish, for example, which is popular in Japan, must have its poisonous gallbladder carefully removed by licensed handlers before it can be sold.

Food and Fish and Shellfish

The texture, taste, and color of fish and shellfish can be affected by their diet. Farmed fish, particularly those in accelerating temperature regimes, can be fatty. Excess fat, however, can be removed by greatly reducing the diet before harvest, and by increasing water circulation to increase energy utilization.

The direct absorption of geosmin from the water can affect the taste of the flesh. Geosmin is produced by some species of actinomycetes and cyanobacteria, which may bloom under certain chemical and physical conditions. They can produce the earthy-muddy flavors that are common in catfish and other freshwater pond fish and are now reported in some marine shrimps. But such flavors can be avoided if stocks are maintained in high-quality geosmin-free water conditions before harvesting. The color of the flesh is readily changed by additives in the diet. Salmonids, for example, are often fed carotenoid pigments prior to harvest to redden the flesh for increased marketing appeal.

The majority of captive aquatic animals are almost entirely dependent on artificially prepared feeds, although they also consume any natural foods that might be available in their enclosures. Artificial diets are invariably high in animal protein (15 to 40 percent, depending on the age and specific needs of the population), complemented by cereal proteins and carbohydrates, oils, and additives of minerals and vitamins.

In recent years the trend has been to formulate diets for fish containing only oils of polyunsaturated fats, and to include chemical attractants and growth promoters. Fish nutrition remains a primary area of research at present because of its direct influence on the quality of the finished product and because feed is the largest operational cost of farming.

The State of Aquaculture in the Present Day

Aquaculture Products on Today's Markets

The general global prosperity of the 1980s increased market demand for a variety of fish and fisheries products, which made the culture of several new species economically feasible. In addition, such demand stimulated production of those species traditionally associated with farming, such as trout, Chinese and Indian carp, and oysters and mussels.

Investments have also continued in the enterprises of raising aquatic animals for recreational use (sport fishing), for medicines and medical research, for biological assay, and for their many valuable by-products. Moreover, many species are cultured successfully for the ornamental fish trade, and in the last 10 years, some new and unusual species have been cultured.

According to the FAO, there are some 150 individual species or species groups under some form of domestic culture for human consumption worldwide, together with 60 more that are identified only by genera and family names. There are about 90 known species of fish currently farmed, 23 species of crustaceans, 35 species of mollusks, 4 species of algae, and miscellaneous aquatic animals (such as frogs, turtles, sea squirts, pearl oysters, and sponges). Of these, some 34 species may be considered the most important, as they are farmed over a wide geographic area, and their individual total production is more than 20,000 tonnes annually.

In Asia, and to a much lesser extent in Eastern Europe, freshwater fishes are the most popular farmed commodities among consumers. The most important are those that have traditional value and have been in great demand for centuries, particularly silver carp, bighead carp, and common carp.

In western Asia, the indigenous Indian carps remain traditional and important. About 7 million tonnes of Chinese and Indian carps are raised annually. They are relatively cheap because they are easily farmed and require few costly inputs. They are usually sold fresh and uncleaned, or "in the round," and many buyers prefer them still alive. The milkfish remains an important staple fish on traditional markets in the Philippines and Indonesia, but production has decreased as coastal ponds have been converted for raising high-value marine shrimp. Because of the bony nature of milkfish, the flesh is removed and reprocessed, but in recent years there has been an increase in the popularity of small kippered "boneless" fish.

In Africa, and to a lesser extent parts of Asia and Latin America, the tilapias are a cheap and widely available farmed commodity, as they are easily raised by small-scale rural farmers in rain-fed ponds. The Nile tilapia is the most popular species because of its larger size. A domestic hybrid, the red tilapia, is proving to be better for small-scale entrepreneurial farming in these regions and is capturing small but

medium-value markets, even in the United States and Europe. Smaller species of tilapia raised in some African countries are used to concoct "relishes," to make leafy energy foods more palatable.

The diverse family of freshwater catfishes all have good medium value. Both local and institutional markets have been developed in the last decade of the twentieth century for these higher-priced species, particularly channel catfish in the United States and South America. Production of channel catfish is so reliable that there is a continuous availability of many product forms for consumers, including "fast foods."

In Europe and North America, the high-value salmonids, which have now been economically important for more than a century, continue to be in great demand. They are raised in both fresh and marine waters. The most popular product for the retail and restaurant trade is pan-size rainbow trout, but by the end of the twentieth century there was intense production of Atlantic salmon, even in Chile and New Zealand. Because of oversupply, prices have dropped considerably, and producers are seeking other markets with a variety of product forms, such as smoked salmon, in a variety of sizes.

The national markets of most developed countries have unmet demands for all farmed marine fish. Compared with freshwater fish, all marine species have medium-to-high value and most are economic to raise. The most popular species currently farmed are mullet, sea bass, bream, grouper, and many marine flatfishes, but progress in raising the volume remains slow.

The most remarkable achievement has been the massive growth in the production of marine shrimp to meet a continuously increasing demand for crustaceans as a whole by markets in Europe, Japan, and the United States. Some 17 of these high-value species are under culture, with the giant tiger prawn and fleshy prawn the most important. Farmed shrimp industries are now economically important to many Asian countries, particularly Indonesia, the Philippines, Thailand, and Taiwan, but they are also important to several countries of Central America and to Ecuador in South America.

Farmed shrimp have advantages in uniformity and, in addition to supplying a range of markets for size, they also satisfy markets for a variety of other product forms. Nonetheless, freshwater crustaceans have not benefited from the popularity of marine shrimps. Farmed species, such as the high-value giant river prawn, freshwater lobsters, and crayfish, are still confined to local markets and usually sold fresh.

Marine mollusks continue to have worldwide consumer appeal. They appear on the luxury markets of developed countries, yet are still farmed for subsistence in some of the poorest countries of Africa. Intensive farm production systems for both oysters and mussels have replaced traditional culture practices in European and Mediterranean countries and

also in North America. In Asia, farmed production of the abalone, Japanese or Manila clam, and the blood cockle has been substantially increased by new and rapidly expanding culture fisheries.

Most developed countries have small luxury markets for a number of specialized aquatic groups. These include the meats of certain farmed reptiles, especially crocodilians and turtles (which also have valuable by-products, such as skins and shells), and of amphibians, particularly frogs, which are raised for their legs.

Summary

Aquaculture has proved to be a remarkable sector of economic growth in recent years. The 1992 value of cultured aquatic animals and plants for food was more than 32.5 billion U.S. dollars, almost triple the first estimate of 12 billion U.S. dollars in 1984. This is due mostly to rapid increases in farm production of aquatic animals, as well as rising prices. Since 1984, farm production of aquatic animals alone has increased by 90 percent. Individually, the production of crustaceans has increased by more than 250 percent, finfish by 105 percent, and mollusks by 55 percent.

Domesticated aquatic animal production contributes significantly to meeting an increasing worldwide food demand. Indeed, annual world demand for fishery products by the year 2000 had been predicted to be 110 to 120 million tonnes, of which about 30 percent is for industrial uses, such as oils and animal feeds. Farming of aquatic animals now contributes about 14 million tonnes annually, almost entirely for human consumption. Because most of these products are sold fresh or iced on local markets, such domestic production supplies about half the world's consumption of fresh fishery products.

In terms of quantity and distribution, the traditional species remain the most important. Like agriculture, aquaculture relies on a small number of "crops" for the greater part of its total production. Farming is still dominated by freshwater fish, particularly the cyprinids (which account for half the total aquatic animal production worldwide), followed by the salmonids and tilapines. The penaeids are the most important family of crustaceans, and mytilids and ostreids are the principal mollusks.

The narrow reliance on a few key crops has implications for the future of aquatic farming, particularly with the increasing involvement of bioengineering in the genetics of these species. On the one hand, there is a need to maintain genetic diversity for the viability of major food crops and the conservation of genetic resources. On the other hand, there is a pressing need for the domestication of farmed aquatic animals, similar to that of terrestrial animals and poultry. Most aquatic animals now farmed are still only one

or two generations removed from natural (or wild) stocks.

Although there is great optimism for the future domestication of aquatic animals through genetic selection and other technical improvements, there are increasing pressures on the industry. In particular, these are the competition for suitable land and water with urban centers and agriculture activities; the impact of nutrient loading on the environment from intensive production; the spread of farm diseases to wild stocks; and the interbreeding of farmed stocks with wild stocks. Nonetheless, despite these growing issues, and a prolonged and shaky history stretching back over 2,000 years, the domestication of aquatic animals seems here to stay.

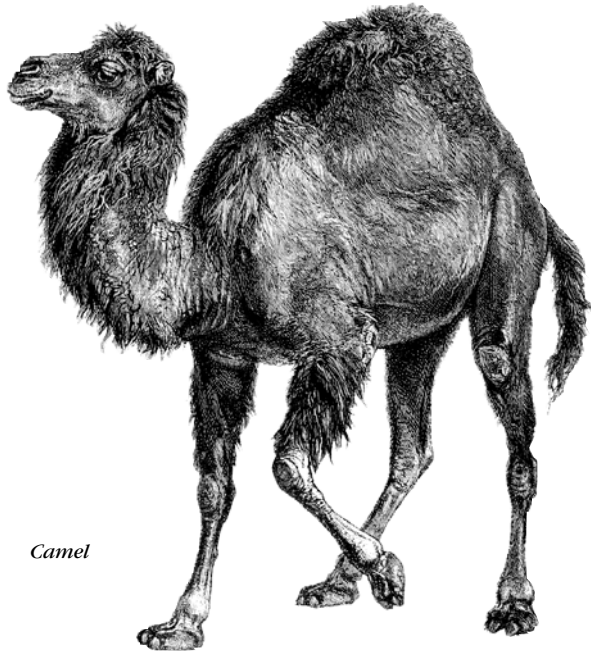
Colin E. Nash

Bibliography

- Chen, T. P. 1976. *Aquaculture practices in Taiwan*. Oxford. China Fisheries Society. 1986. *Fan Li on pisciculture*. Peking.
- Dobrai, L., and G. Pékh. n.d. *Fisheries in Hungary*. Budapest.
- Dunfield, R. W. 1985. *The Atlantic salmon in the history of North America*. Canadian Special Publications of Fisheries and Aquatic Sciences 80. Ottawa.
- Dyk, V., and R. Berka. 1988. Major stages of development in Bohemian fish pond management. *Papers of the Research Institute of Fishery and Hydrobiology, Vodňany, Czechoslovakia* 17: 3-44.
- Graham, M. 1956. *Sea fisheries*. London.
- Hepher, B., and Y. Pruginin. 1981. *Commercial fish farming: With special reference to fish culture in Israel*. New York.
- Hickling, C. F. 1962. *Fish culture*. London.
- Hora, S. L., and T. V. R. Pillay. 1962. *Handbook on fish culture in the Indo-Pacific region*. Rome.
- Matsui, I. 1979. *Theory and practice of eel culture*. Tokyo.
- Radcliffe, W. 1926. *Fishing from the earliest times*. London.
- Shelbourne, J. E. 1964. The artificial propagation of marine fish. *Advances in Marine Biology* 2: 1-83.
- Toussaint-Samat, Maguelonne. 1992. *History of food*, trans. Anthea Bell. Cambridge, Mass.
- Vibert, R., and K. F. Lagler. 1961. *Pêches continentales*. Paris.

II.G.3 Camels

Camels are familiar to most of the English-speaking world only as beasts of burden, as Arab mounts on the silver screen, or as curiosities in the zoo. Camel meat and milk almost never find their way to the local grocer's shelves. In many parts of Africa and Asia, however, camel milk and meat are valuable sources of nutrition for many people. The following is a broad overview of the camel's current role as a food resource and a summary of evidence dealing with the early history of human reliance on camel products.



Camel

The people who keep camels for food are generally nomadic pastoralists, who rely primarily on livestock for their livelihood and move regularly in order to care for their animals, which, in addition to camels, may include other species such as sheep, goats, and cattle. In both Africa and Asia, camel pastoralists inhabit arid environments characterized by extreme, often unpredictable, fluctuations in temperature and rainfall. The

latter may be restricted to one or two short seasons, may fail to happen in any given year, or may be localized in a very small geographic area, leaving vast stretches of the environment parched and barren. As a result, the productive potential of these ecosystems varies greatly over time and space, and those inhabiting such areas must be able to adapt quickly to variability in food, forage, and water availability. Camels are often a vital part of the pastoral strategy for coping with such harsh conditions. They cover great distances in search of limited and highly scattered forage and water, which is then efficiently converted into milk and meat for human consumption. In times of prolonged heat or drought, camels may be the only productive livestock and, consequently, are invaluable resources during critical months or seasons of scarcity.

Two Camel Species

There are two extant species of camel: the one-humped dromedary (*Camelus dromedarius*) of the hot deserts of Africa, Arabia, and India; and the two-humped bactrian (*Camelus bactrianus*) of the seasonally cold, higher deserts of the Iranian plateau, central Asia, China, and Mongolia (Figures II.G.3.1 and II.G.3.2). The camel's closest living relatives belong to the once species-rich suborder Tylopoda, which today includes the llama, alpaca, guanaco, and vicuña of South America. In this chapter, the generic term "camel" is used to refer to both species of Old World camels when it is not necessary to distinguish between them. However, it is the case that dromedaries and bactrians are uniquely adapted to the extremes of their native environments, resulting in important differences between them.

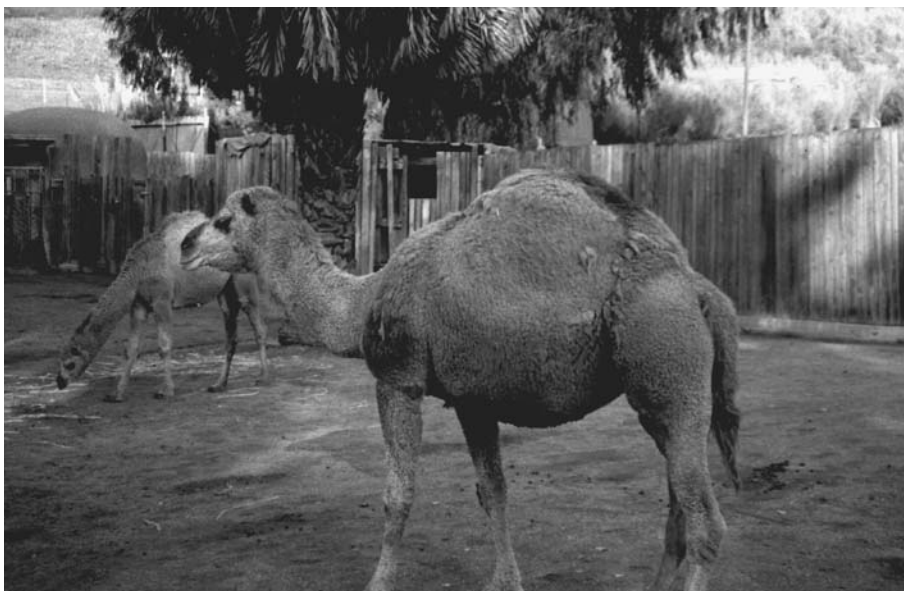
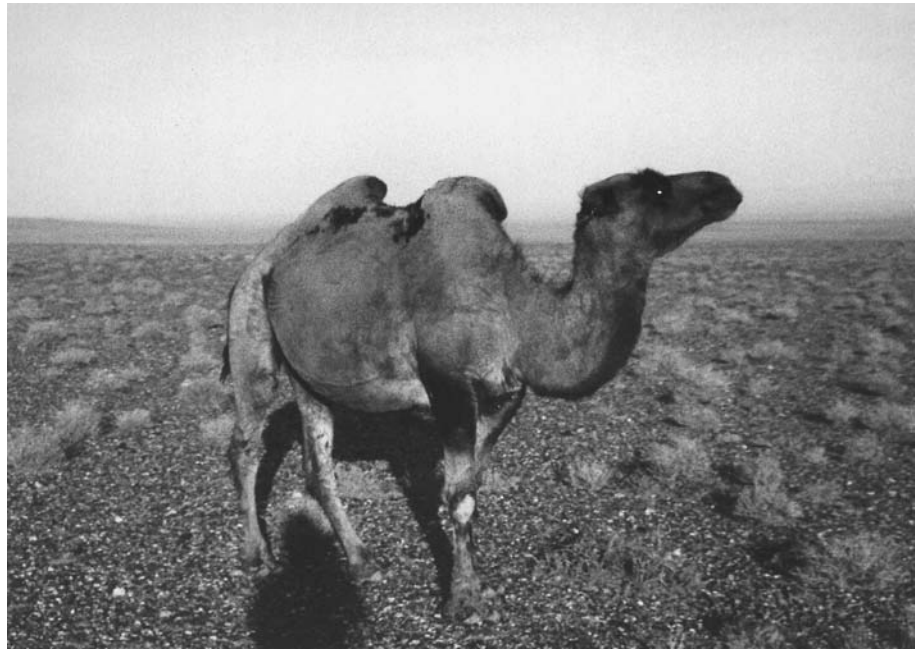


Figure II.G.3.1. *Camelus dromedarius*.

Figure II.G.3.2. *Camelus bactrianus*.



Dromedaries have longer legs, their winter coats are much thinner, and their summer coats are more reflective than those of bactrians. These traits are suited to surviving the intense and prolonged heat of the Arabian and African summers. Bactrians are also adapted to surviving in arid conditions, but their stockier builds and thick winter coats allow them to weather the icy winters of the highland deserts of Asia. In China, bactrians regularly experience temperatures of -30°C (-22°F) in the winter and 42°C (107.6°F) in the summer (Wei 1984). Although bactrians can survive in a wide range of temperatures, they are usually not found where the mean annual temperature is above 21°C (69.8°F) (Mason 1984).

Dromedaries and bactrians are not perfect biological species because they can (and do) mate. The resulting hybrids are larger than either parent and have one elongated hump. Unlike the hybrid offspring of horses and donkeys, first-generation camel hybrids are fertile (Gray 1954). However, if two first-generation hybrids are mated, the resulting second-generation hybrids display suboptimal qualities and are often infertile (Gray 1954; Bulliet 1975). Hybrid camels can be successfully mated with either a pure-blooded dromedary or a pure-blooded bactrian. The resulting offspring resemble the pure-blooded parent in size and number of humps (Bulliet 1975).

Historically, these first-generation hybrid camels were popular caravan animals due to their larger size and greater stamina than either purebred parent (Bulliet 1975; Toplyn 1995). Hybrid camels were also better able to survive the climatic extremes encountered at either end of the trade routes that connected China

and the Iranian plateau with Arabia and Egypt. In order to maintain a constant supply of these large-bodied hybrids for caravan traffic, pastoralists had to continuously breed pure dromedary and bactrian parents so as to obtain viable hybrids. Today, hybrid camels are still being produced in the former Soviet Union, where dromedaries of the Turkmen tribes are bred with Kirghiz bactrian camels (Bulliet 1975).

Geographic Distributions

It is not clear from archaeological and paleontological data whether the natural geographic ranges of dromedaries and bactrians overlap or whether their modern ranges are the result of human intervention. Like horses, camels originated in North America and migrated to Asia around 3 to 4 million years ago (Wilson 1984; Stanley, Kadwell, and Wheeler 1994). During the Pleistocene, many species of camel inhabited the Eurasian and African landmasses. Their bones have been recovered across Asia from Siberia to India, from Eastern Europe, and across North Africa as far south as Olduvai Gorge (Howell, Fichter, and Wolff 1969; Badam 1979; Wilson 1984; Dubrovo and Nigarov 1990; Germonpré and Lbova 1996).

Indeed, camels seem to have been successful animals that enjoyed a wide distribution until the beginning of the Holocene (about 10,000 years ago), after which their remains become extremely rare in the archaeological record of central and southwestern Asia (Bulliet 1975; Zarins 1978; Hoch 1979; Wapnish 1984; Uerpman 1987; Tchnernov 1989; Germonpré and Lbova 1996). The few pieces of evidence from the early Holocene suggest that wild camels were

confined to the southern part of the Arabian Peninsula and the deserts of central Asia, and it was not until about 1500 B.C. that domestic camels began to spread across southwestern and central Asia to occupy their current ranges.

In other words, prior to domestication, wild camels seem to have been close to extinction – inhabiting severely restricted ranges in southern Arabia and in parts of central Asia (Bulliet 1975; Russell 1988). But today, domestic camels inhabit the arid regions of North and East Africa, the Arabian Peninsula, parts of Israel, Jordan, Syria, Turkey, Iraq, Iran, Pakistan, Afghanistan, northwestern India, central Asia, northwestern China, and southern Mongolia (Map II.G.3.1). Dromedaries were also introduced into Australia in the late 1800s, and today 15,000 to 20,000 feral dromedaries roam the central and western parts of that continent (McKnight 1969).

Access to water seems to have played a pivotal role in the decline of wild camel populations and

may have been a significant factor in their domestication. Despite the camel's ability to survive without drinking for extended periods (weeks to months depending on the water content of their forage), eventually they do have to replace their water losses. Studies of wild camels in China and Mongolia indicate that severe droughts force them to congregate around the few remaining sources of water, which, among other things, exposes them to increased predation by wolves (Tulgat and Schaller 1992). In fact, recent droughts have caused a sharp decline in the wild camel populations in these areas. Yet domestic camels seem to avoid the problems of drought and predation through their association with humans. In Africa, Arabia, and Mongolia, camels often rely on humans to access water in wells or cisterns, especially when rain-fed waterholes dry up (Yagil 1985). Some camels have been known to wait by a well for weeks in order to secure water.



Map II.G.3.1. The approximate modern distribution of camels.

Camels, however, cannot always afford to flee a drought-stricken area for greener pastures and more abundant water because moist environments can also be deadly, especially to dromedaries. Humid environments with standing water are often breeding grounds for tsetse flies that transmit trypanosomiasis and other diseases (Fratkin and Smith 1994), which kill many camels every year and limit the distribution of the dromedary in Africa and India (Wilson 1984; Yagil 1985). Pastoralists are keenly aware that certain swampy areas spread disease and will take their camels there for water only when no other options are available (Fratkin and Smith 1994).

In addition, camels are notoriously maladapted to wet, muddy ground, which, because of their unique physiology of long legs, unilateral gait, and large padded feet, makes them unstable. They often pull muscles, dislocate joints, or break limbs if forced to cross such areas (Russell 1988). Moreover, camels are "front-wheel-drive" animals. Unlike horses, camels use their forelimbs rather than their hind limbs for propulsion. This trait, in combination with padded rather than hooved feet, leaves them at a severe disadvantage in rocky and mountainous areas (Sato 1980; Russell 1988).

Food and Water Requirements

Compared to other common domestic livestock, such as cattle, sheep, and goats, camels are the most efficient users of water, much of which they can obtain from their herbivorous diet. When necessary, they conserve water by producing a very concentrated urine and dry feces and by reducing water lost through perspiration. In response to solar radiation, dromedaries tolerate core body temperature fluctuations from 34° C to 42° C (93.2° F to 107.6° F) throughout a day (Schmidt-Neilsen 1964), thereby decreasing the need to sweat to regulate body temperature. They can also tolerate a loss of significantly more water relative to their body weight than other livestock, which reach critical limits when they lose 10 percent of their body water (Russell 1988). Dromedaries regularly lose one-third of their body weight in water and survive (Yagil 1985).

The camel's humps are fat deposits that also aid in water conservation. By concentrating fat along the back, the rest of the body is able to cool itself more efficiently. The dorsal fat concentrations not only are insulation against the brunt of solar radiation but also allow the body to radiate heat more efficiently than would be possible if fat were distributed evenly over the whole body. In addition, behavioral mechanisms aid in cooling. During the hottest part of the day, dromedaries tend to lie near each other, facing the sun, which minimizes body surface exposure to direct solar radiation. Calluses on the knees and chest of the camel allow it to rest on hot sands and to elevate its midsection, which enhances airflow beneath the body. Thus, the animal can reach a

greater cooling potential. When water is available, a thirsty camel can replace even extensive water deficits by drinking as much as 200 liters in 3 minutes (Yagil 1985).

The camel's efficient use of water enables it to access seasonally and spatially dispersed high-quality forage that other livestock cannot. Often, forage protein content is inversely related to water availability (Russell 1988). Forage that is abundant after periodic rainfall or in areas around predictable water sources generally contains less protein than plants grown in drier areas. Camels move almost continuously as they graze, enabling them to cover many kilometers (km) each day; in fact, they can graze from 50 to 80 km away from water sources, whereas sheep and goats are limited to 20 to 30 km and cattle to 4 to 13 km (Dahl and Hjort 1976; Sato 1980; Russell 1988). As a result, camels are able to use the more concentrated packages of plant protein that are scattered across the landscape and located far enough away from water sources to be inaccessible to other livestock.

Camels are primarily browsers and are able to eat plant species that other livestock cannot, such as camel thorn, *Acacia* species, and saltbushes (Yagil 1982). When camels, cattle, sheep, and goats are herded together, they compete for some of the same plants, especially grasses (Sato 1980); consequently, camels are usually herded separately to ensure adequate forage for all livestock.

Although all livestock need salt in their diet, camels need more than sheep, goats, and cattle, and to meet these requirements, they must regularly visit areas with saline water or salt-rich plants (Sato 1980; Bernus 1990; Jabbur 1995). Alternatively, pastoralists may collect salt crystals or salt-rich earth to carry back to their camel herds.

In short, camels are well adapted to living in arid environments because they have evolved many mechanisms to conserve water and are able to consume a wide range of plant species. These adaptations decrease competition with other herbivores that share the same environments. From a husbandry standpoint, however, such adaptations often create extra work for pastoralists. If camels must be watered from a well, the human labor needed to satisfy their thirsts can be substantial, as even a small herd of 10 to 20 animals may need thousands of liters of water (Yagil 1985). This probably explains why, in the biblical story of Rebekah, Abraham's servant was so impressed by Rebekah's offer to water his 10 camels (Gen. 24:29).

Camels' Productive Potential: Labor and Food

Camels are physiologically much better adapted than humans to living in the deserts and semi-deserts of Africa and Asia, yet many people inhabit arid regions of the world. As part of their survival skills, humans have learned to use products from

camels and to harness their power. However, not all people utilize camels and camel products the same way.

Labor

Historically, camels were very important for carrying loads over long distances and difficult terrain, and early long-distance trade flourished because of camel caravans. Exotic spices were transported from southern Arabia to the urban centers of Egypt and Mesopotamia during the first millennium B.C. Later, in the early centuries A.D., camels carried trade items along the Silk Route between Persia and China (Bulliet 1975). As mounts and water carriers, camels were also valuable in desert warfare. Some of the earliest literary accounts that mention domestic camels tell of mounted pastoralists from southern Arabia raiding villages to the north, and the animals continued as military mounts until as recently as the early part of the twentieth century in Arabia and India (Bulliet 1975). Camels have also served less dramatic but important roles in the daily lives of nomadic pastoralists, and still today, nomadic pastoralists use the animals to carry dwellings, household belongings, and old or very young people during regular migrations and relocations (Hobbs 1989).

With minimal loads, dromedaries can maintain a pace of 16 km (10 miles) per hour for up to 18 hours (Yagil 1985). Depending on the age, size, and health of the camel, a load of over 300 kilograms (kg) can be carried for several days, and up to 800 kg can be carried for short trips (Gauthier-Pilters and Dagg 1981). Bactrians in China continue to be used primarily as beasts of burden that can easily carry normal loads, averaging 40 percent of their live weight, 35 to 40 km in 24 hours (Wei 1984). Riding camels are also common in China and Mongolia; in fact, in the Chinese deserts, a bactrian camel is considered a more comfortable mount than a horse, and it maintains a pace of 5 to 10 km per hour (Wei 1984).

Today, however, the role of the camel as a transport animal is declining in many parts of the world. Except where automobiles and gasoline are still prohibitively expensive, camels are being replaced by motorized transport.

Milk Quality, Quantity, and Processing

Cross-culturally, milk is the most important edible product obtained from camels. Whether consumed fresh, processed for storage, or mixed with the milk of other livestock, this nutritious product is harvested by almost all pastoralists who keep camels. The absolute quantity and specific qualities of each camel's milk varies with the quantity and quality of available forage and water and the animal's health and age. The nutritional value of camel milk varies because of cultural norms regulating its collection, processing, and distribution within and outside of the pastoral group.

Milk quality: According to J. S. Jabbur (1995), camel milk is higher in vitamin C, fat, protein, and minerals than milk from other livestock raised under the same desert conditions. Certainly, this source of vitamin C is very important to people who inhabit areas where fresh fruits and vegetables are not available year-round. Camel milk contains between 5.7 and 9.8 mg of vitamin C, which is 3 times the amount found in cow's milk and 1.5 times the amount found in human milk (Yagil 1982).

The water content of camel milk varies inversely with the amount of available drinking water. Studies have found that when water was freely accessible, camels produced relatively concentrated milk, consisting of around 86 percent water; when drinking water was restricted, however, the milk became more diluted, with a water content of some 91 percent (Yagil 1982; el Amin 1984; Knoess 1984; Shalash 1984b). Such an adaptation is extremely valuable to both camel calves and people because, in conditions of heat stress and mild dehydration, the diluted milk serves as a source of needed liquid.

The fat content of camel milk also changes in response to the state of hydration of the animal. A hydrated dromedary produces milk with 3.0 percent to 5.38 percent fat, whereas the milk of one that is dehydrated may contain only 1.1 percent fat (Yagil 1982; el Amin 1984; Knoess 1984). Bactrian milkfat has been measured at slightly higher levels – 5.8 percent to 6.6 percent (Shalash 1984b). The various milks produced by other livestock species contain similar amounts of fat. Water buffalo, cows, goats, and sheep produce milk with 7.5, 3.5, 2.8, and 3.7 percent fat, respectively (Knoess 1984). Structurally, camel milkfat is somewhat different from cow, sheep, and goat milkfat in that it is distributed in very small globules (1.2 to 4.2 microns), which are also bound to proteins (Yagil 1982). This structure makes it very difficult to separate the fat from the milk and probably explains claims that butter cannot be made from camel milk. In fact, camel milk can be made into butter and cheese, but with some considerable difficulty, and consequently, when available, sheep, goats, and cattle are generally the preferred sources of milk for these products.

Lactose levels and protein concentrations do not vary with the hydration level of the animal. There are, however, individual and breed differences. Protein may constitute between 2 and 5.5 percent of camel milk, and some camels produce sweeter-tasting milk with lactose levels of 5.8 percent or more. Lower lactose levels are also common and give the milk a slightly sour taste (Yagil 1982). Buffalo, cow, goat, and sheep milk contain similar levels of lactose (Knoess 1984), although those of horse and human milk are higher (6.9 and 7.0 percent, respectively) (Shalash 1984b).

Milk quantity: It is difficult to compare milk yields between dromedaries and bactrians and between different populations of camels from around the world.

Daily milk yields seem to vary a great deal with the season, the age of the calf, and the availability of forage and water. Lactation begins when a camel gives birth, and its duration varies depending on food and water availability. Milk production will normally cease when the camel becomes pregnant again or if the calf dies, but pastoralists have developed ingenious methods of encouraging the camel to continue to produce milk when the latter occurs. For example, the camel is often tricked into thinking that her calf is still alive by the pastoralists' covering another calf with the skin and scent of the dead calf and encouraging it to suckle (Yagil 1982).

As a rule, bactrians lactate continuously for 14 to 16 months (Wei 1984), whereas dromedaries lactate for 12 to 20 months after a calf is born (Sato 1980; Yagil 1982). In parts of Africa, sheep and goats will give milk for only 3 to 4 months, and cattle for about 10 months (Sato 1980). Therefore, if people relied only on cattle or sheep and goats for milk, they would have to do without an important part of their diet for several months at a time.

Various methods of measuring camel milk yields have generated a wide range of daily and annual production estimates. Most of these are not explicit about whether the volume of milk obtained represents only the amount destined for human consumption or whether it also includes the amount normally consumed by the calf. Nevertheless, a survey of the literature indicates that dromedaries produce more milk on average than bactrians and that dromedary milk yields from across Asia, India, and Africa vary between 3.5 and 50 kg per day (Yagil 1982; Knoess 1984). In East Africa, camels produce an average of 3.5 to 4.0 liters of milk per day in the dry season and up to 10 liters per day in the wet season (Fratkin and Smith 1994). This quantity is quite impressive when compared to cattle in the same environment, which were able to produce only 1 liter per day in the wet season and 0 to 0.2 liters in the dry season. In northern Kenya, camels give twice as much milk as cattle and nine times that of sheep and goats (Sato 1980).

A camel's ability to produce milk is also influenced by the number of times it is milked per day, with more frequent milkings increasing a dam's total milk yield by 10 to 12 percent (Shalash 1984b). Milking frequency is a function of the herding practices employed and the household's need for milk. If camels must graze far away from the household, then they are milked less frequently than animals that graze nearby. In Sudan, hydrated camels yield 2.5 to 5 kg of milk at midday, and morning and evening milkings each yield 3.5 to 5 kg, which means a daily milk yield of 8.5 to 15 kg (Hjort and Dahl 1984). Raikas in India can obtain up to 10 liters of milk per day from their camels by milking them three to five times daily (Köhler-Rollefson 1992).

Bactrian camels produce an average of 5 kg of milk

per day, although 15 to 20 kg have been obtained from specific animals. In China, most of the milk is left for the calf, and only about 2 kg is milked for human consumption (Yagil 1982).

Milk processing. Cultural practices can have a large effect on the nutritional benefits of camel milk, and many differences exist in milk-processing methods, the percentage of the diet derived from milk and milk products, and the distribution of the milk products within and outside the group. In most instances, camel milk is mixed with the milk of other lactating animals before it is consumed or processed, and if the milk is not to be consumed fresh, it is quickly processed to keep it from spoiling. Such processing usually results in those fermented milk products that are ubiquitous wherever milk is consumed. They are known by a variety of names around the world – such as kefir, *matzoon*, *dabdi*, yoghurt, and *lehben*, to name a few – but all are produced by the same basic method.

First, the milk is boiled for several minutes, which kills any pathogenic bacteria that may be in it. Once the milk has cooled to 79° F to 86° F, a starter culture from a previous batch of fermented milk is added. The milk is stirred and allowed to ferment overnight at room temperature, after which the resulting product can be safely stored for several days at room temperature (Yagil 1982). The fermentation process generally decreases the lactose content of the milk by converting it to either ethyl alcohol or lactic acid depending on the type of bacteria used. In human populations that do not maintain high lactase (the enzyme needed to digest lactose) levels as adults, soured or fermented milk is the preferred form for consumption (Simoons 1981). However, most human populations that have kept camels for centuries maintain high lactase levels as adults (Yagil 1982).

Another common method of preserving raw milk is to gently boil and evaporate it over a fire. The resulting solids (fats and proteins) are mixed with sugar to form a sweet, buttery, semisolid, cheese-like substance (Yagil 1982). To make butter requires up to 4 hours of churning or the use of a special blender, and the result generally has a greasy consistency. Most often, it is used for cooking, as a cosmetic, and for medicinal purposes. Buttermilk, left over from the butter-making process, often serves as a base for soup (Yagil 1982).

Milk as a Dietary Staple

G. Dahl and A. Hjort (1984) have calculated that 4 kg of camel milk will deliver an individual's daily caloric needs and more than enough protein, which means that a herd of 18 to 20 camels can supply the nutritional needs of a six-person family (two adults and four children). Most camel pastoralists, however, do not live on milk alone; domestic grains, hunted game, and gathered wild plants also contribute significantly to their diets (Hobbs 1989; Fratkin and Smith 1994; Jabbur 1995).

Cultural practices surrounding the use of camel milk vary greatly around the world. In India, for example, its consumption is limited to the camel-breeding caste known as the Raikas (Köhler-Rollefson 1996), and both the sale and the processing of camel milk into curds is taboo, although it may be given away (Köhler-Rollefson 1992). Raikas usually drink camel milk fresh, but on some occasions it is boiled and sweetened or made into rice pudding. Such practices and regulations ensure that most of the milk produced by Indian camels is not consumed by humans and that few Indian people outside of the Raika caste ever taste it.

In Arabia, Africa, and central Asia, however, the camel's milk-producing ability is fully utilized. The Rwala of Arabia claim (apparently correctly) that camel milk, supplemented with wild game, roots, and seeds, constitutes a sufficient diet (Lancaster and Lancaster 1990), although camel milk has recently become less important in the diets of Arabian Bedouins, who are being forced to settle permanently and sell their camel herds (Köhler-Rollefson 1993). For many in Africa, however, milk – including that of camels – is still a major dietary component. It supplies fully 75 percent of the daily calories in Rendille diets, 60 percent in Maasai diets, 60 to 70 percent in Boran diets, and 62 percent in Turkana diets (Fratkin and Roth 1990). In Kazakhstan, milk and milk products can supply up to 90 percent of the daily diet, and although this milk comes from several species of livestock, camels provide 37 percent of the total, sheep another 30 percent, yak 23 percent, and cows 10 percent (Yagil 1982).

Meat

Worldwide, labor and milk are the most important products obtained from camels, and often, camel pastoralists must choose husbandry strategies that optimize the production of milk and labor at the expense of meat yields. Such strategies emphasize the retention of adult females and the culling of young, unnecessary males, with labor supplied by nonpregnant females and/or castrated adult males (Wapnish 1984). The result is to limit the number of mature animals that can be harvested for meat. Generally, only old animals and young males not needed for breeding purposes are slaughtered. In many traditional societies, young camels are slaughtered only for special ceremonies and celebrations, and other unwanted animals are usually sold in commercial markets (Wapnish 1981; el Amin 1984; Köhler-Rollefson 1996).

In modern Sudan, however, pastoralists who raise camels rarely consume their meat but rather sell old animals to international markets for consumption in Egypt, Libya, and Saudi Arabia (el Amin 1984). Meat from old camels is tough and (save for their livers) not highly regarded among urban consumers, whereas meat from camels 5 years old and younger is

as tender (and as highly valued) as beef; in fact, camel meat is said to taste like coarse, slightly sweet beef (el Amin 1984; Shalash 1984a).

The slaughter of an adult camel produces a large quantity of meat. Male dromedary carcasses may weigh more than 400 kg, whereas a bactrian may exceed 650 kg (Yagil 1982; Wei 1984). In general, a dressed dromedary carcass yields between 52.6 and 76.6 percent meat (el Amin 1984), and dressed bactrian carcasses have been reported to yield 35.4 to 51.7 percent (Wei 1984). The meat is generally very lean, with the total fat content of a carcass varying between 0 and 4.8 percent, depending on the age and nutritional status of the animal. The fat is concentrated in the hump and around the kidneys; that from the hump is melted and used instead of butter for cooking by some Arabian Bedouins (Jabbur 1995).

In addition to husbandry considerations, cultural and religious proscriptions limit the use of camel meat in many countries. In India, for example, Hindus will not eat camel meat, and it is also avoided by Christian Copts of Egypt, Christian Ethiopians, Zoroastrians of Iran, Mandaeans of Iraq and Iran, Nosaioiris of Syria, and many Jews (Yagil 1982).

Blood

Several groups in eastern Africa use the blood of camels (as well as that of cattle) as a food resource that supplies valuable nutrients, including iron, salts, protein, and vitamin D (Yagil 1982). Blood may be drawn from a camel up to twice a month, and the animal can produce between 2 and 5.5 liters with each bleeding (Shalash 1984a; Bollig 1992). Blood is usually processed before consumption by first collecting it in a bowl and stirring it with a stick. Because fibrin, a blood-clotting protein, collects on the stick, it is easily separated from the rest of the blood and generally roasted before it is eaten. The remaining blood is either mixed with milk or boiled and mixed with maize flour (Bollig 1992).

Other Products

Camels provide many other valuable products, including wool, urine, and dung. In Sudan and India, dromedary wool is collected as the animals begin to molt and is used to make jackets, robes, tents, ropes, blankets, and carpets (el Amin 1984; Köhler-Rollefson 1992). In China and Mongolia, bactrian wool is a highly valued product that is often made into cushions, mattresses, bags, and ropes (Wei 1984). Among the Bedouins of Arabia, camel urine is used as a hair tonic (Jabbur 1995). Apparently its high salt content gives hair a reddish tint and prevents vermin infestations. Dung, also a valuable commodity, is often dried and stored for heating and cooking fuel (Jabbur 1995). And finally, if a person is lost in the desert and desperate for water, a camel may be slaughtered for the water contained in its stomach (Jabbur 1995).

Husbandry Considerations

Given the previous discussion of the camel's productive potential, one might wonder why camels are not more widely and intensively herded, or why people bother to herd other livestock with their camels. These questions are best answered by examining the camel's reproductive characteristics and the resulting husbandry strategies employed by camel pastoralists. Pastoralists have developed a variety of strategies that optimize the camel's productive potential in their particular social and ecological environment. Specifically (as already noted), pastoralists must balance the harvesting of camel products today against the need for future herd growth. This is usually calculated in terms of how much milk and meat can be taken for human consumption versus the amount of milk required by calves and the number of animals necessary to meet future production needs and to hedge against future unpredictable calamities.

Life History and Reproductive Characteristics

An animal's ability to reproduce quickly is a valuable trait to pastoralists. Reproductive potential depends on the age of first reproduction, the potential number of offspring per year, and infant mortality. A quick comparison of annual calving rates – 0.4 or less for camels, 0.4 to 0.8 for cattle, and 0.8 to 2.0 for sheep and goats – shows that camels are among the slowest reproducers of any domestic livestock species (Russell 1988; Herren 1992). Their low rate results from a 12- to 13-month gestation period (which is shorter for dromedaries than bactrians) and, typically, a 24- to 36-month span between births. Other factors that lower reproductive potential include the frequent restriction of breeding to rainy seasons and susceptibility to trypanosome, brucellosis, and pasteurellosis infections that trigger abortions.

The interval between births can be narrowed with intensive management practices, but this practice appears to shorten the female's life and limit the total number of offspring (Bollig 1992). Because camel herds increase at a rate of 8 percent or less annually, it takes a minimum of 15 years, and potentially as many as 50 years or more, for one to double in size (Dahl and Hjort 1976; Russell 1988; McCabe 1990). Such slow reproductive rates directly affect how people manage camel herds for food production and how they manage social obligations and their own reproductive rates (Galvin, Coppock, and Leslie 1994). Because camels will lactate only after giving birth and usually stop lactating shortly after becoming pregnant again, pastoralists carefully manage herds to ensure that at least some animals are lactating at any given time. But even with good luck and careful management, only 15 to 22 percent of the females may be lactating at one time. The camel's low calving rate also limits the number of young males that can be slaughtered for meat (Herren 1992). But the trade-off is that

if a calf is slaughtered, all of the dam's milk becomes available for human consumption, providing, of course, that the dam can be tricked into continuing lactation.

Human Labor

Human-labor requirements are an important obstacle to the efficient utilization of domestic animals in pastoral societies because the availability of this labor limits the total number of livestock that can be herded by a single household. As we have seen, differing needs for food, water, preferred forage, rates of locomotion, duration of milk production, and breeding cycles prevent sheep, goats, cattle, and camels from being herded in a single group. Labor requirements are not constant throughout the year, but as herds are divided into milking and nonmilking and breeding and nonbreeding groups, more labor is required (Sato 1980; Roth 1990; Fratkin and Smith 1994). Pastoralists also face the constant problem of maintaining that delicate balance between labor requirements necessary to care for livestock and the number of people that can be supported by a given number of livestock (Fratkin and Smith 1994). To achieve and maintain this balance was doubtless at the root of the development of many inheritance, exchange, ritual, and human population control practices among camel pastoralists (Sato 1980; Fratkin and Smith 1994; Galvin et al. 1994).

Differences in livestock water requirements, forage plant species preferences, and grazing speed frequently force pastoral subsistence groups to split up for large portions of the year (Sato 1980; Sperling 1987; Bollig 1992; Fratkin and Smith 1994). This split often takes place along age or gender lines, resulting in dramatically different diets among members of the society (Sato 1980; Galvin et al. 1994). Illustrative are the Rendille of Kenya, who strictly delegate performance of the various herding tasks. Camels may be herded and milked only by young unmarried men; consequently, this segment of the Rendille population spends the longest time away from the main camp and covers the greatest distances during the grazing rounds. While on herding duty, these men subsist solely on camel milk and blood. Unmarried men and women herd small stock (sheep and goats) and consume both maize meal and milk while away from the base camp during the dry season. Adults, children, and the elderly, who remain at the base camp throughout the dry season, must subsist primarily on maize meal because few lactating animals can be supported on the meager forage available near the camp (Sato 1980).

Mixed Herding Strategies

Given that herding several different livestock species requires the division of pastoral societies and more herding labor than "camel-only" strategies, one might ask why so many pastoral societies go to such trou-

ble. The answer is that a mixed pastoral system is an effective hedge against episodes of food scarcity. For example, if a camel herd is wiped out by a disease, it may take a lifetime to rebuild it. Alternatively, sheep and goats, although more likely to die during extended droughts, are able to quickly reproduce their numbers. Moreover, in most pastoral economies, sheep and goats satisfy immediate needs for meat and are often exchanged for grains.

Camels, however, are kept for the products that can be obtained from live animals during times of scarcity. In short, pastoralists use sheep and goats as we would use spending money, whereas camels are a form of stored wealth that might be equated with long-term stock or bond investments. In addition to maintaining diverse livestock herds, many pastoralists further diversify their resource base through trading relationships. They exchange animals and animal products for agricultural and industrial products and may also "loan" animals to relatives and neighbors or "adopt" children if herding requirements exceed the available labor (Russell 1988). These relationships can be used to optimize herd production and labor utilization over a greater geographic area.

Moreover, camels and other livestock can be viewed as a form of stored food for pastoralists. They not only represent food reserves that can be utilized on a regular basis but also serve to concentrate scattered or inedible nutrients into highly nutritious and accessible packages for human consumption (Russell 1988). Additionally, to return to the theme of stored wealth, the ability of camels to survive severe droughts better than other livestock, as well as their 30-year lifespan, makes them desirable "banks." For example, during a series of droughts in northern Africa between 1968 and 1973, the Tuareg lost 63 percent of their cattle and 47 percent of their sheep but only 38 percent of their camels (Bernus 1990). After a severe drought, camels can be sold or exchanged for necessities and to replace lost livestock.

The Camel in Antiquity

Very little is known about the earliest culinary uses of camels because of the nature of the archaeological record and because of the camel's preference for desert environments. The archaeological record is generally best preserved within structures such as buildings, cities, and natural enclosures such as caves. Camels, however, do not, and presumably did not, live in these environments. Not surprisingly, then, very few archaeological sites have produced evidence for the early use of camels. Before they were domesticated, camels were hunted for their meat, hides, and bones, and it is doubtful that they were exploited for their milk until after domestication.

But before humans could hunt wild camels or, later, milk domestic camels, humans first had to come into contact with them, and human-camel interaction

is rarely documented in the archaeological record prior to the third millennium B.C. What little evidence of such contact is available comes from southwestern Asia (the presumed refuge of the wild dromedary) and (for the bactrian camel) from sites on the Iranian plateau, in central Asia, and in China (Map II.G.3.2).

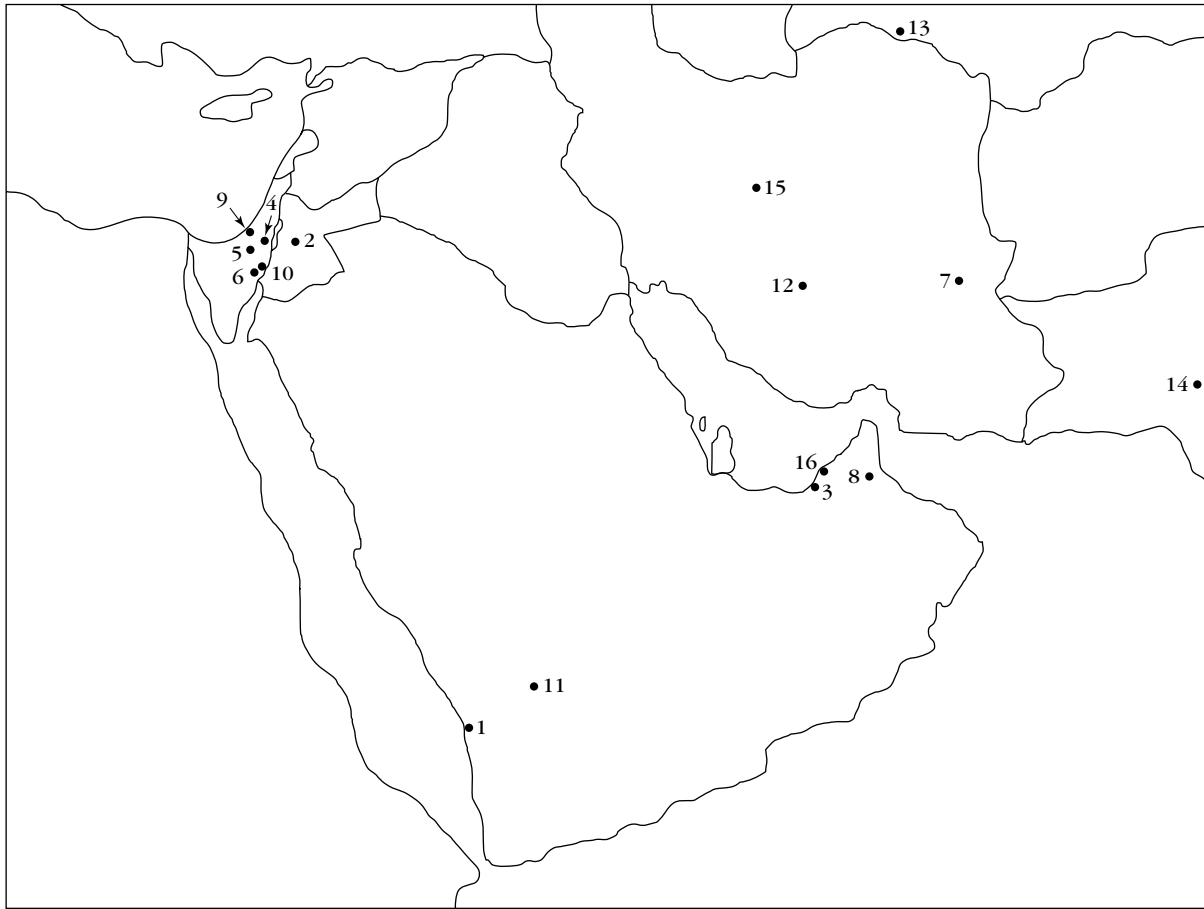
Dromedary Domestication

Camel bones from Paleolithic sites (before 10,000 B.C.) in Egypt, Sudan, Jordan, Syria, and Israel attest to the wide distribution of a camel species – presumably the dromedary – prior to the Holocene. Subsequently, however, camel remains become much more rare, suggesting that they either were not in the vicinity of archaeological sites or, for some unknown reason, were not a regular food source.

Some of the earliest rock art, dated between 6000 and 3500 B.C., shows dromedaries and speared dromedaries (Anati 1970; Zarins 1989), and archaeological evidence confirms the presence of camels in southern Arabia at this time. At Sihi, a shell midden on the Red Sea coast, a camel mandible was recovered and directly radiocarbon-dated to about 7000 B.C. (Grigson 1989). Moreover, camel bones from southern Jordan were found in seventh-millennium-B.C. contexts (Köhler 1984), whereas other camel remains discovered in the United Arab Emirates were likely from the fourth millennium B.C. (Uerpmann 1987).

In southern Arabia, rock art images showing hunting scenes, hunting parties, and the use of bows and arrows provide pictographic evidence that local hunters were responsible for the camel bones found at several of the area's sites and dating from between 3500 and 1900 B.C. (Zarins 1989). Third-millennium-B.C. sites along the Gulf coast of the United Arab Emirates have also yielded numerous such bones (Hoch 1979; Uerpmann 1987), including those from the island site of Umm an-Nar, where, although it is not clear whether the camels were wild or domesticated, they were most certainly introduced by humans (Hoch 1979).

The long history of human-camel interaction in southern Arabia indicates a likely environment for the occurrence of domestication but does not document exactly when such an event might have happened. However, it would seem that we can actually spy early domestic camels to the north of Arabia. In the Sinai and the Levant, camel images are absent from rock art traditions (suggesting that wild camels were unknown), but their bones nonetheless begin to appear in second-millennium-B.C. urban sites (Lernau 1978; Hakker-Orion 1984; Zarins 1989). For an explanation, we return to southern Arabia, where people for millennia had kept sheep, goats, and cattle and lived in permanent structures made of stone. But then, presumably, the ecological changes caused by desiccation and a concomitant expansion of the desert forced the adoption of a more mobile lifestyle. It is believed, in other words, that the bones found on



Map II.G.3.2. Archaeological sites mentioned in the text.

- | | | | |
|-----------------|------------------|-------------------------|------------------|
| 1. Sihi | 5. Bir Resisim | 9. Tell Jemmeh | 13. Anau |
| 2. Ain al Assad | 6. Timna | 10. Wadi Arabah site #2 | 14. Mohenjo-daro |
| 3. Umm an-Nar | 7. Shar-i Sokhta | 11. Bir Hima region | 15. Sialk |
| 4. Arad | 8. Hili 8 | 12. Tale Malyan | 16. Ras Ghanada |

the northern edge of the desert were those of domesticated camels, upon which the now mobile pastoralists from southern Arabia had begun to visit the urban centers of the north (Tchernov 1989; Zarins 1989).

Bactrian Domestication

The early history of human-bactrian interaction is poorly understood at present. This is partly because of a limited amount of archaeological investigation of central Asian sites and partly because of the inaccessibility of most Russian and Chinese documents to outside researchers. Although evidence for the use of bactrian camels in Neolithic China is practically nonexistent (because few animal bones were ever saved or studied), an exception is one camel bone from a site near Lake Burkol in the northeastern part of Xinjiang Province; this bone has been radiocarbon-

dated to 3000 B.C. (Olsen 1988). Rock art from northern China and Inner Mongolia also attests to the presence of bactrians; however, dates for this art are highly speculative. Not until the Zhou Dynasty (c. 1100–771 B.C.) do Chinese literary sources confirm that bactrian camels were domesticated and employed as beasts of burden (Olsen 1988).

Early evidence for the presence of camels on the Iranian plateau is found at the site of Shar-i Sokhta (Compagnoni and Tosi 1978), where bones, dung (contained in a ceramic jar), and hair (woven into a piece of fabric), dating from roughly 2600 B.C., were found. These discoveries suggest that people were already utilizing several valuable resources from camels that presumably lived near the site. Other finds indicate that camels in the Indus Valley at the site of Mohenjo-daro by 2300 B.C. (Meadow 1984)

were bactrians. A pottery shard from a contemporary site on the Iranian plateau depicted a bactrian, and no images of dromedaries have been found from this period (Compagnoni and Tosi 1978). However, there is no consensus on whether the bones from these areas are those of dromedaries or bactrians (Compagnoni and Tosi 1978; Meadow 1984; Wapnish 1984), largely because there is evidence of extensive trading between the Arabian Peninsula and Iran by the first half of the third millennium B.C. (Hoch 1979), and dromedaries could have been introduced to the Iranian sites from Arabia the trading network.

Domestication as Beasts of Burden

The distribution of camel bones at Tell Jemmeh, an urban site in southern Israel, may reveal when and why camels finally became an important domestic animal across southwestern Asia. Their bones are absent from the earliest levels of this site – the Chalcolithic (c. 3200 B.C.) and Middle Bronze II (c. 1650–1550 B.C.) (Wapnish 1981), with the first deposits of camel remains dating from between approximately 1400 and 800 B.C. But over this 600-year period, only 7 camel bones were left behind to be recovered. Eight more showed up for the following century, when the Assyrians began to exert their influence at the site through military campaigns, and 40 bones were deposited between 675 and 600 B.C. – a 75-year period that corresponds to the city's association with the Assyrian invasions of Egypt. Then, from the beginning of the sixth century B.C. until occupation of the site ended at around 200 B.C., over 273 camel bones were left behind. These last four centuries witnessed periods of political instability as first neo-Babylonians, then Persians, and finally Alexander the Great took control of the site.

The question of why camel remains steadily increased in this urban center – if it was well supplied with meat from domestic cattle, donkeys, sheep, goats, pigs, and wild game – may be answered by the fact that camels found increasing economic and political importance as beasts of burden. The long-distance trade in incense and spices from southern Arabia to the Levant and beyond had become lucrative. Empires wanted to control and profit from this trade but needed to be able to supply and move troops in these arid regions. Both enterprises depended on the camel's unique ability to haul heavy loads over long distances while consuming little water, and a whole economy grew up around this ability (Wapnish 1981). Archaeological investigations at two other sites in the region also support this interpretation. Camel bones from Wadi Arabah No. 2 and Timna, two copper-production sites (1350–1150 B.C.), indicate that camels were used to transport this valuable product to urban markets (Rothenberg 1972; Zarins 1989).

At the site of Tal-e Malyan, in present-day Iran, a similar pattern of camel introduction is evident. Although the site was occupied from at least 3400

B.C. and has yielded thousands of animal bones, those of camels do not appear until the Middle Elamite period (1600–1000 B.C.) (Zeder 1984). Regional political instability may have led to the city's decline during these centuries, encouraging people to abandon their urban lifestyles for a more mobile pastoral existence. Camels introduced to the region by caravaneers figured prominently in the new pastoral system.

The Assyrians were instrumental in introducing camels to people across Southwest Asia, although not always peacefully, and their texts and monuments provide us with much of the available evidence about the historical use and distribution of dromedaries and bactrians. Some of the earliest written accounts of dromedaries come from an Assyrian cuneiform text, written between 1074 and 1057 B.C., which mentions that the animals were brought to Nineveh and exhibited with other exotic animals (Zarins 1989). Two bactrian camels are depicted on the Black Obelisk of Shalmaneser III of Assyria, which has been dated to 856 B.C. (Bulliet 1975).

All of the preceding tends to fit rather nicely with a synthetic hypothesis of the origins of dromedary pastoralism proposed by J. Zarins (1989). The camel was probably hunted for its meat in southern Arabia until around 2200 B.C., when dromedaries were domesticated and kept for milking purposes. However, they were not ridden or used as beasts of burden until around 1500 B.C. – about the time when saddles were developed for them. Afterward, camels were increasingly employed for riding and as pack animals for overland trade. After about 1000 B.C., saddle technology was improved, and camels became very important in warfare as mounts for armed soldiers (Bulliet 1975).

The origin of domesticated bactrians is more obscure. Presumably, the process of their domestication followed a similar path, although we do not know if they were domesticated earlier or around the same time as dromedaries. It is clear, however, that by the first millennium B.C., if not earlier, bactrians, too, were being used as pack animals (Bulliet 1975).

The archaeological record also plainly indicates that overland trade and warfare introduced camels to peoples who lived beyond the animals' native deserts. Apparently, these people appreciated the camel's ability to produce copious amounts of milk year-round and to carry large burdens over great distances in environments that were inhospitable to most other livestock. Moreover, the late appearance of indisputably domesticated camels suggests that they were incorporated into already-established pastoral lifestyles with sheep, goats, and cattle.

Our understanding of the origins of the use of camels as a food resource should improve as current efforts in identifying and studying the archaeological traces of early pastoralists are completed and published. In addition, ethnographic studies of modern camel pastoralists that attempt to quantify the costs

and benefits of camel husbandry under specified ecological and social conditions would enhance our understanding of the origins of such practices. And finally, an ongoing study of camels' genetic variability may someday help to trace the origins and subsequent dispersal of camels and the pastoral systems associated with camel husbandry.

Elizabeth A. Stephens

A Jacob Javits Fellowship and the Haury Educational Fund for Archaeology generously funded research leading to the production of this chapter. Fieldwork was graciously supported by Dr. John Olsen. I would also like to thank Dr. Mary Stiner, Dr. Carol Kramer, Dr. Stephen Zegura, Dr. Michael Hammer, Jeff Brantingham, Lane Therrell, Dr. Andrew Standeven, and anonymous reviewers for their help, encouragement, and valuable comments on my research.

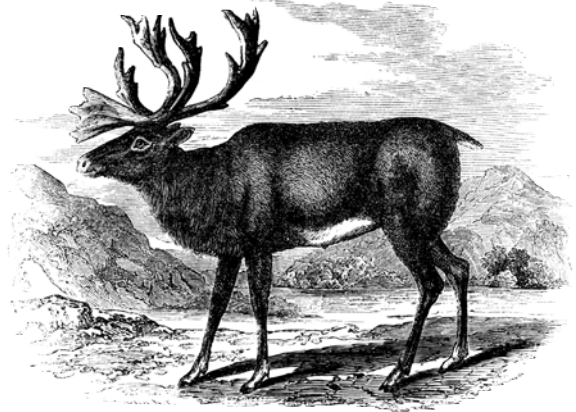
Bibliography

- Anati, E. 1970. Rock engravings of Dahthami wells in central Arabia. *Bollettino del Centro Camuno di Studi Preistorici* 5: 99-158.
- Badam, G. L. 1979. *Pleistocene fauna of India: With special reference to Siwaliks*. Deccan College, India.
- Bernus, E. 1990. Dates, dromedaries, and drought: Diversification in Tuareg pastoral systems. In *The world of pastoralism: Herding systems in comparative perspective*, ed. J. G. Galaty and D. L. Johnson, 149-76. New York.
- Bollig, M. 1992. East Pokot camel husbandry. *Nomadic Peoples* 31: 34-50.
- Bulliet, R. W. 1975. *The camel and the wheel*. Cambridge, Mass.
- Cashdan, E. A. 1985. Coping with risk: Reciprocity among the Basarwa of northern Botswana. *Man* 20: 454-74.
- Compagnoni, B., and M. Tosi. 1978. The distribution and state of domestication in the Middle East during the third millennium B.C. in light of finds from Sharh-I Sokhta. In *Approaches to faunal analysis in the Middle East*, ed. R. H. Meadow and M. A. Zeder, 91-103. Cambridge, Mass.
- Dahl, G., and A. Hjort. 1976. *Having herds: Pastoral herd growth and household economy*. Stockholm.
1984. Dromedary pastoralism in Africa and Arabia. In *The camelid: An all-purpose animal*, Vol. 1, ed. W. R. Cockrill, 144-66. Uppsala, Sweden.
- Dubrovo, I. A., and A. N. Nigarov. 1990. Plio-Pleistocene fossil vertebrate localities of south-western Turkmenia, U.S.S.R. *Quartarpalaontologie* 8: 35-45.
- el Amin, F. M. 1984. The dromedary of Sudan. In *The camelid: An all-purpose animal*, Vol. 1, ed. W. R. Cockrill, 36-49. Uppsala, Sweden.
- Fratkin, E., and E. A. Roth. 1990. Drought and economic differentiation among Ariaal pastoralists of Kenya. *Human Ecology* 18: 385-402.
- Fratkin, E., and K. Smith. 1994. Labor, livestock, and land: The organization of pastoral production. In *African pastoralist systems: An integrated approach*, ed. E. Fratkin, K. A. Galvin, and E. A. Roth, 91-112. Boulder, Colo.
- Galvin, K. A., D. L. Coppock, and P. W. Leslie. 1994. Diet, nutrition, and the pastoral strategy. In *African pastoralist systems: An integrated approach*, ed. E. Fratkin, K. A. Galvin, and E. A. Roth, 113-204. Boulder, Colo.
- Gauthier-Pilters, H., and A. I. Dagg. 1981. *The camel: Its evolution, ecology, behavior, and relationship to man*. Chicago.
- Germonpré, M., and L. Lbova. 1996. Mammalian remains from the Upper Palaeolithic site of Kamenka, Buryatia (Siberia). *Journal of Archaeological Science* 23: 35-57.
- Gray, A. P. 1954. *Mammalian hybrids: A check-list with bibliography*. Bucks, England.
- Grigson, C. 1989. The camel in Arabia - A direct radiocarbon date, calibrated to about 7000 B.C. *Journal of Archaeological Science* 16: 355-62.
- Hakker-Orion, D. 1984. The role of the camel in Israel's early history. In *Animals and archaeology*, Vol. 3, *Early herders and their flocks*, ed. J. Clutton-Brock and C. Grigson, 207-12. Oxford.
- Hartly, B. J. 1984. The dromedary in the Horn of Africa. In *The camelid: An all-purpose animal*, Vol. 1, ed. W. R. Cockrill, 77-97. Uppsala, Sweden.
- Herren, U. J. 1992. Cash from camel milk: The impact of commercial camel milk sales on Garre and Gaaljacel camel pastoralism in southern Somalia. *Nomadic Peoples* 30: 97-113.
- Hjort, A., and G. Dahl. 1984. A note on the camels of the Amar'ar Beja. In *The camelid: An all-purpose animal*, Vol. 1, ed. W. R. Cockrill, 50-76. Uppsala, Sweden.
- Hobbs, J. J. 1989. *Bedouin life in the Egyptian wilderness*. Austin, Tex.
- Hoch, E. 1979. Reflections on prehistoric life at Umm an-Nar (Trucial Oman), based on faunal remains from the third millennium B.C. *South Asian Archaeology 1977*: 589-638.
- Howell, F. C., L. S. Fichter, and R. Wolff. 1969. Fossil camels in the Omo Beds, southern Ethiopia. *Nature* 223: 150-2.
- Ingold, T. 1983. The significance of storage in hunting societies. *Man* 18: 553-71.
- Jabbur, J. S. 1995. *The Bedouins and the desert: Aspects of nomadic life in the Arab East*, trans. L. I. Conrad. Albany, N.Y.
- Knoess, K. H. 1984. The milk dromedary. In *The camelid: An all-purpose animal*, Vol. 1, ed. W. R. Cockrill, 176-95. Uppsala, Sweden.
- Köhler, I. 1984. The dromedary in modern pastoral societies and implications for its process of domestication. In *Animals and archaeology*, Vol. 3, *Early herders and their flocks*, ed. J. Clutton-Brock and C. Grigson, 201-6. Oxford.
- Köhler-Rollefson, I. 1992. The Raika dromedary breeders of Rajasthan: A pastoral system in crisis. *Nomadic Peoples* 30: 74-83.
1993. Camels and camel pastoralism in Arabia. *Biblical Archaeologist* 56: 180-8.
1996. The one-humped camel in Asia: Origin, utilization and mechanisms of dispersal. In *Origins and spread of agriculture and pastoralism in Eurasia*, ed. D. R. Harris, 282-95. Washington, D.C.
- Lancaster, W., and F. Lancaster. 1990. Desert devices: The pastoral system of the Rwala Bedu. In *The world of pastoralism: Herding systems in comparative perspective*, ed. J. G. Galaty and D. L. Johnson, 177-94. New York.
- Lernau, H. 1978. Faunal remains, strata III-I. In *Early Arad*, ed. R. Amiran, 83-113. Jerusalem.

- Mason, I. L. 1984. Origin, evolution, and distribution of domestic camels. In *The camelid: An all-purpose animal*, Vol. 1, ed. W. R. Cockrill, 16-35. Uppsala, Sweden.
- McCabe, J. T. 1990. Turkana pastoralism: A case against the tragedy of the commons. *Human Ecology* 18: 81-103.
- McKnight, T. L. 1969. *The camel in Australia*. Melbourne.
- Meadow, R. H. 1984. Animal domestication in the Middle East: A view from the eastern margin. In *Animals and archaeology*, Vol. 3, *Early herders and their flocks*, ed. J. Clutton-Brock and C. Grigson, 309-27. Oxford.
- Newman, D. M. R. 1984. The feeds and feeding habits of Old and New World camels. In *The camelid: An all-purpose animal*, Vol. 1, ed. W. R. Cockrill, 250-92. Uppsala, Sweden.
- Olsen, S. J. 1988. The camel in ancient China and an osteology of the camel. *Proceedings of the Academy of Natural Sciences of Philadelphia* 140: 18-58.
- Roth, E. A. 1990. Modeling Rendille household herd composition. *Human Ecology* 18: 441-5.
- Rothenberg, B. 1972. *Were these King Solomon's mines?* New York.
- Russell, K. W. 1988. *After Eden: The behavioral ecology of early food production in the Near East and North Africa*. Oxford.
- Sato, S. 1980. Pastoral movements and the subsistence unit of the Rendille of northern Kenya, with special reference to camel ecology. *Senri Ethnological Studies* 6: 1-78.
- Schaller, G. 1995. Tracking the Gobi's last wild bears and camels. *International Wildlife* 25: 18-23.
- Schmidt-Neilsen, K. 1964. *Desert animals: Physiological problems of heat and water*. Oxford.
- Shalash, M. R. 1984a. The production and utilization of camel meat. In *The camelid: An all-purpose animal*, Vol. 1, ed. W. R. Cockrill, 231-49. Uppsala, Sweden.
- 1984b. The production and utilization of camel milk. In *The camelid: An all-purpose animal*, Vol. 1, ed. W. R. Cockrill, 196-208. Uppsala, Sweden.
- Simoons, F. J. 1981. Geographic patterns of primary adult lactose malabsorption. In *Lactose digestion: Clinical and nutritional implications*, ed. D. M. Paige and T. M. Bayles, 23-48. Baltimore, Md.
- Sperling, L. 1987. The adoption of camels by Samburu cattle herders. *Nomadic Peoples* 23: 1-18.
- Stanley, H. F., M. Kadwell, and J. C. Wheeler. 1994. Molecular evolution of the family Camelidae: A mitochondrial DNA study. *Proceedings of the Royal Society of London* 256: 1-6.
- Tchernov, E. 1989. Faunal turnover and extinction rate in the Levant. In *Quaternary extinctions: A prehistoric revolution*, ed. P. S. Martin and R. G. Klein, 528-52. Tucson, Ariz.
- Toplyn, M. R. 1995. Meat for Mars: Livestock, *limitanei*, and pastoral provisioning for the Roman army on the Arabian frontier, A.D. 284-551. Ph.D. thesis, Harvard University.
- Tulgat, R., and G. B. Schaller. 1992. Status and distribution of wild bactrian camels: *Camelus bactrianus ferus*. *Biological Conservation* 62: 11-19.
- Uerpmann, H.-P. 1987. *The ancient distribution of ungulate mammals in the Middle East*. Wiesbaden.
- Wapnish, P. 1981. Camel caravans and camel pastoralists at Tell Jemmeh. *Journal of the Ancient Near Eastern Society of Columbia University (JANES)* 13: 101-21.
1984. The dromedary and bactrian camel in Levantine historical settings: The evidence from Tell Jemmeh. In *Animals and archaeology*, Vol. 3, *Early herders and their flocks*, ed. J. Clutton-Brock and C. Grigson, 171-200. Oxford.
- Wei, D. 1984. The bactrian camel of China. In *The camelid: An all-purpose animal*, Vol. 1, ed. W. R. Cockrill, 98-111. Uppsala, Sweden.
- Wilson, R. T. 1984. *The camel*. London.
- Yagil, R. 1982. *Camels and camel milk*. Food and Agriculture Organization of the United Nations, Animal Production Paper No. 26. Rome.
1985. *The desert camel*. Basel.
- Zarins, J. 1978. The camel in ancient Arabia: A further note. *Antiquity* 52: 44-6.
1989. Pastoralism in Southwest Asia: The second millennium B.C. In *The walking larder*, ed. J. Clutton-Brock, 127-55. London.
- Zeder, M. A. 1984. Meat distribution at the highland Iranian urban center of Tal-e Malyan. In *Animals and archaeology*, Vol. 3, *Early herders and their flocks*, ed. J. Clutton-Brock and C. Grigson, 279-307. Oxford.

II.G.4 🦌 Caribou and Reindeer

The terms "caribou" and "reindeer" refer to a species of cervid, *Rangifer tarandus*, which has Holarctic distribution. "Reindeer," however, is somewhat ambiguous, as it is used to refer to both the wild and domesticated forms of this species, whereas "caribou" always designates the wild form. Yet "reindeer" is generally preferred in the Old World for both. (The term "wild reindeer" is also sometimes used as a synonym for "caribou," to differentiate from the domesticated animal.) In addition, a separate species of Paleoarctic (Old World) reindeer, *Rangifer arcticus*, was once recognized, but the fact that caribou and reindeer interbreed successfully and produce fertile offspring, particularly in Alaskan herds, led to general unification of the taxon (Banfield 1961).



Reindeer

At least six major modern subspecies of caribou are recognized. Two of these are Paleoarctic in distribution: *Rangifer tarandus tarandus*, the European tundra reindeer, and *Rangifer tarandus fennicus*, the Eurasian forest reindeer. Three are Neoarctic in distribution: *Rangifer tarandus grant* of Arctic Alaska and westernmost Canada, *Rangifer tarandus pearyi* (the "Peary caribou") of eastern Arctic Canada, and *Rangifer tarandus groenlandicus* of Greenland and Baffin Island (Meldgaard 1986). One subspecies, *Rangifer tarandus caribou*, is endemic to the subarctic boreal forest, originally ranging from Alaska to eastern Canada and the northernmost United States but now more limited in distribution. Other varieties are more restricted geographically: *Rangifer tarandus platyrhynchus* lives only on Spitzbergen, and *Rangifer tarandus dawsoni* is confined to the Queen Charlotte Islands.

Prehistory of Caribou/Wild Reindeer Utilization

The genus *Rangifer* exists in the paleontological record throughout the middle to late Pleistocene or Ice Age period and has a probable antiquity of at least 400,000 years. The relative prevalence of caribou remains in paleontological and archaeological sites from this period serves as a sensitive indicator of climatic and vegetational change (Parker 1971; Messier et al. 1988). This is because caribou are tied to arctic tundra or subarctic taiga (open coniferous forest) habitats, which provide their chief foods: lichens, particularly of the genera *Cladonia* and *Alectoria*, and a wide variety of low browse or groundcover plants, including forbs, fungi, willow and birch shoots, and grass and sedge shoots (Kelsall 1968). Thus, the presence of caribou during glacial periods (and their absence during interglacial periods) in central Europe and other parts of northern Eurasia is linked to the expansion and regression of the Scandinavian ice sheet to which these arctic and subarctic environments were tied (Bouchud 1966).

Remains of *R. tarandus* are primarily associated with late *Homo erectus* or archaic *Homo sapiens* sites in Europe and, more frequently, with Neanderthal sites. *Rangifer tarandus* existed as a recognizable taxon throughout the Upper Paleolithic period from 40,000 to 10,000 years ago (Banfield 1961). It is a conservative taxon, with little change occurring in caribou throughout the late Ice Age and postglacial periods. During the Upper Paleolithic period, many northern Eurasian hunting populations became specialized in hunting wild reindeer (Sturdy 1975; Mellars 1996; Burke and Pike-Tay 1997). The caribou was probably the most important game animal in western Europe during the Upper Paleolithic, and European cave deposits contain large amounts of caribou bones (Delpech and Heintz 1976; Delpech 1983; Boyle 1990, 1993; Straus 1997; Thacker 1997). Caribou also figure prominently in Ice Age art, as at Lascaux Cave, and

some depictions (as at Trois Frères) suggest that they played an important role in ritual and perhaps shamanistic activities. In the New World, caribou were important to Paleo-Indian hunters throughout North America (Cleland 1965; Funk, Fisher, and Reilly 1970; Spiess, Curran, and Grimes 1985; Peers 1986; Jackson 1988).

After the end of the Ice Age, caribou populations retreated northward to areas offering suitable habitat, such as northern Germany and southern Scandinavia, and many parts of Russia, Alaska, Canada, and Greenland. They were an important focus of Epipaleolithic or Mesolithic hunters throughout this region; their bones are found in archaeological sites, and they were depicted in rock art in each of these areas. With evolving post-Pleistocene climates, caribou became restricted to areas of northern Scandinavia, Siberia, parts of the Russian Far East, and the more northerly areas of Canada and Alaska (Yesner 1995). They have continued to be hunted by various aboriginal populations in all of these regions until the present time (e.g., Birket-Smith 1929; Lips 1947; Chard 1963; Gubser 1965; Nellenman 1970; Simchenko 1976; Irimoto 1981; Hall 1989; Krupnik 1993).

Methods of Caribou/Wild Reindeer Exploitation

Since Upper Paleolithic times, caribou have been a historically important food resource for northern Eurasian hunting peoples for three reasons: the dense aggregation of the animals in bands and herds, their ease of capture, and their nutritional value. Modern caribou tend to congregate in large numbers ranging from dozens (bands) to thousands (herds). Their migratory habits minimize their impact on the relatively delicate plant communities upon which they depend. These migrations may shift groups between forested and open (tundra) environments on a seasonal basis. They frequently involve movement between seasonally important hunting grounds and spring "calving grounds," where females give birth. Contemporary migration distances range from relatively short (for example, on Spitzbergen and islands of the Canadian High Arctic) to very long (in western Canada and Alaska). Factors involved in the distance and location of migration include available forage, protection against predators, and insect infestations (Pruitt 1960; Skoog 1968; Heard 1997).

The nature of caribou migration patterns of the past is unclear. Zooarchaeological data from late Ice Age sites in southwestern France suggest that year-round occupation of some areas was possible by specialized caribou hunters, which may be related to shorter-distance migrations by the animals, as well as to the fact that they existed in significant numbers (Gordon 1988; David and Enloe 1992; Enloe and David 1997; Spiess in press). In comparison with their Arctic habitats of today, southwestern France would have provided a smaller physical area but much

greater environmental (particularly altitudinal) diversity, thereby helping to support larger groups of animals. In addition, beginning around 20,000 years ago, caribou were hunted with efficient tools such as spears and atlatls (Bouchud 1966; Delpech 1983; Pike-Tay and Bricker 1993).

In places where they were less available, such as in open forest areas, caribou were a seasonally important resource; their dietary significance seems to have varied in relation to the specificity of the habitat for caribou and the length of local caribou migration routes. Among historically known caribou hunting groups, some maintained a herd-following strategy, whereas others remained in specific areas, usually near annual migration routes, exploiting other resources until the caribou showed up.

The former pattern seems to have been more typical of places where few alternative resources existed, such as the Canadian High Arctic (Gordon 1975), and may also have been true of Scandinavia before the domestication of reindeer (Blehr 1990). The second pattern was more typical of places where alternative resources existed, where annual migration routes were fairly precise, and where caribou could be taken in sufficiently large numbers that their meat could be stored for a considerable time period; such a location was the Brooks Range of northern Alaska (Burch 1972). Situations in which caribou composed the predominant element in the diet yet failed to materialize along a particular migration route in a given year have been well documented; the result was usually starvation for the hunters (Mowat 1952).

Human groups hunting caribou have employed a wide variety of techniques, which Arthur Spiess (1979), Otto Blehr (1990), and Bryan Gordon (1990) have discussed in detail. There are basically two approaches: One involved "mass killing" techniques, in which large numbers of animals were taken from a band or herd at one time, whereas the other saw individual caribou or small numbers of individuals killed. The former tended to be used for hunting large bands or herds (as during seasonal aggregations) when large numbers of humans were available to hunt and where caribou meat was particularly important in the diet. Mass killing techniques involved logistic planning, whereas individual hunting techniques were largely based on opportunistic encounters.

Mass killing techniques were historically used by many groups, including coastal and interior Eskimos, various Athapaskan peoples of interior Alaska and western Canada, some Algonquian groups in the eastern Canadian Subarctic, the Saami peoples of northern Scandinavia, some Siberian groups, and a few of the Rocky Mountain groups of western Canada. Algonquians of northern New England and peoples of the Russian Far East seem to have used individualistic hunting techniques almost entirely. Mass killing techniques were of two types. On land, they involved drives into traps, snares, fences, corrals, or human "sur-

rounds"; in general, this type included any technique in which the caribou were lured into a small area to be killed and butchered (Anell 1969; Morrison 1982). The caribou were also driven into water, with the use of boats, in areas where rivers or lakes lay near migration routes (Arima 1975; Greenman and Stanley 1977; Yesner 1980). Male and female humans of all ages participated in these activities, with which forms of sociopolitical leadership were often associated. These may have ranged from simple task leadership to the involvement of more general leaders or village headmen (for example, the *umialik* or whaling boat captains in North Alaskan villages).

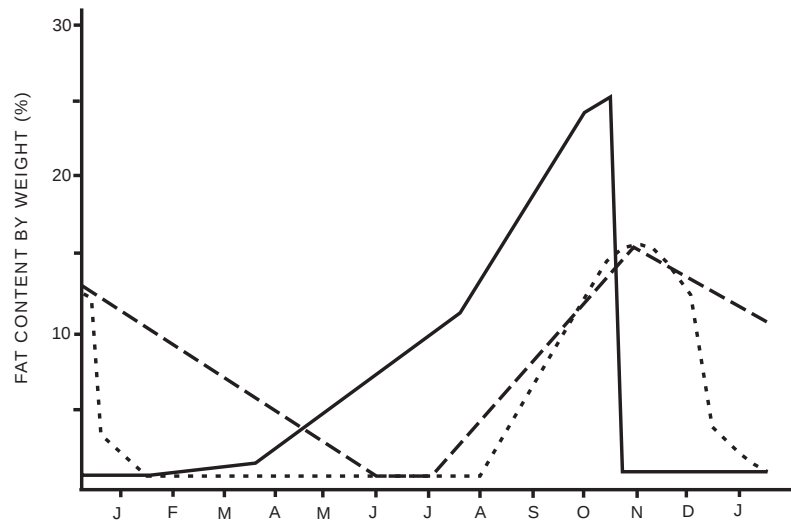
By contrast, individualistic hunting was usually done by men using singly set snares, stalking, and "running down" of game. Killing was done with atlatls, spears, bows and arrows, and the rifles introduced into many indigenous caribou hunting groups during the nineteenth century. For both mass and individualistic hunting various aids were employed, such as decoys (antlers, reindeer skins, calls) and mock caribou made of wood or stone. Some of these approaches may be reflected in the archaeological record; examples might include the caribou head-dress from Trois Frères or certain depictions in Scandinavian rock art. Mass killing efforts tended to yield nonselective age and sex groups of caribou, whereas individual hunting techniques were more effective at different times of year to take particular ages or sexes of the animals. The archaeological record indicates that bulls were generally preferred for meat, and today they are larger and have more fat, particularly during the autumn (Yesner 1980). Calves were also sought for their high fat content, and female adults became the preferred prey in early winter, when the bulls were in rut (see Figure II.G.4.1).

Individualistic hunting techniques were also used to take animals at specific times of the year for non-food purposes, such as in late summer and fall, when the quality of reindeer fur was particularly suitable for clothing manufacture. Cows and calves were often sought for their skins. Animals of varying ages and sexes were targeted for the manufacture of specific items of clothing: skin bags, rawhide or "babiche" for snowshoe lacings, as well as covers for dwellings, caches, and sleds (Smith 1978). In addition, antlers were an important resource for tool manufacture. If already shed, antlers could be scavenged, but they could also be taken from animals, particularly males, and generally from adults over 4 years of age, as they produce significantly larger antlers.

Dietary Importance of Caribou/Wild Reindeer

The fact that specialized caribou hunting has been possible among some groups for 40,000 years indicates that the animal is capable of satisfying human subsistence requirements (Burch 1972). However, rarely did historically known groups subsist solely on

Figure II.G.4.1. Seasonal variation in the fat content of caribou. Mature bulls are represented by a solid line, mature females by a dashed line, and young bulls by a dotted line. (From Spiess 1979: 28; used by permission of Academic Press.)



a caribou meat diet. Past caribou-hunting specialists of the arctic region, such as the Arctic Small Tool peoples (also known as the Denbigh peoples in Alaska and the Dorset peoples in Canada), were historically replaced by more efficient coastal hunters, such as the Thule peoples (ancestral to modern Arctic Eskimos), who combined caribou and sea-mammal hunting. There is also historical and archaeological evidence suggesting that contemporary groups – such as the Nunamiut Eskimos of Alaska, the Caribou Eskimos of Canada, and even the so-called Reindeer Lapps of northern Scandinavia – have developed their specialized life-ways only in the past few hundred years, and that until recently, they migrated between coastal and interior regions, combining caribou hunting with fishing and sea-mammal hunting.

The Caribou Eskimos, living to the west of Hudson Bay in an area where whaling or walrus hunting is less possible today than in other areas (and where shore-fast ice requires individual stalking of seals in winter), have been particularly dependent on caribou. They are, however, a group with smaller human populations, simpler technologies, and less complex political structures than other Eskimo peoples. Similarly, among Alaskan Athapaskan groups, those involved primarily with caribou hunting (for example, the Gwich'in) have been more mobile, with a simpler technology, and have been fewer in numbers than other groups that were also involved with salmon fishing and/or seal hunting. Such caribou-dependent groups tends to be very mobile, occupying in the course of a year several different base camps, where they have utilized relatively unsophisticated dwellings such as caribou-hide tents (or even brush lean-tos).

Widespread concern over the availability of caribou was especially characteristic of groups occupying boreal forest regions, where the animals were often

scarce (cf. Smith 1978). Such concern took the form of specialized ritual and may have stimulated the development of divination as a technique in caribou hunting. The effect of divination was to randomize the impact of such hunting and therefore minimize the likelihood of overexploitation (Moore 1969). Some authors have even linked the so-called Windigo Psychosis of subarctic Cree peoples to concerns over starvation in the boreal forest.

That caribou-hunting groups occasionally met with starvation is well documented in the ethnographic record. In part, this was a result of the naturally cyclical patterns of caribou abundance in arctic regions. Although it is difficult to establish the antiquity of these patterns, there is some suggestion that they go back hundreds of years in interior Alaska (Yesner 1980), Labrador (Fitzhugh 1972), and Greenland (Gronnow, Meldgaard, and Nielsen 1983; Meldgaard 1983), and they may even have occurred during the Upper Paleolithic in France (David 1973). During caribou population highs, such as in late-nineteenth-century Alaska, intensification of hunting, including widespread construction of “caribou fences,” seems to have taken place. At times of population lows, caribou-hunting groups apparently shifted to a more diverse diet that included small game and birds (Yesner 1989). Within the last 150 years, habitat disturbance has led to population expansion of other cervids such as the moose, which subsequently filled the dietary niche formerly occupied by caribou in some parts of interior Alaska.

Because higher metabolic rates are required for human survival under cold conditions, dietary requirements of high-caloric foods are significantly greater in the Arctic region (Draper 1974). Calories are usually provided by fats, which have twice the caloric density of protein or carbohydrates (about 9

calories per gram). However, except for deposits around viscera, under the skin, and in bone marrow, caribou are generally lean, yielding fat at only 10 percent of body weight (Hall 1971). Between 40 and 55 percent of a caribou carcass is meat (White 1953; Hill 1967; Binford 1978). Caribou weights vary from about 35 kilograms (kg) for a juvenile to 150 kg for a large adult male (Kelsall 1968). Such a range provides from 20 to 80 kg of meat but only 3 to 10 kg of fat (Table II.G.4.1).

There is some physical evidence that nutritional deprivation was occasionally associated with those groups that were more caribou-dependent. In the Canadian High Arctic, for example, a greater dependence on caribou may be linked to relatively shorter statures in that region (Laughlin 1966). Although there is evidence for episodic nutritional stress in human skeletons from throughout the Arctic and Subarctic (Yesner 1994), the Caribou Eskimos have shown some of the strongest patterns of development of growth arrest lines, indicative of starvation, of any Eskimo peoples (Buikstra 1976). In addition, studies by the Alaska Dietary Survey (Mann et al. 1962) and by Edward Foulks (1972) have suggested that a caribou diet may be deficient in thiamine and other vitamins.

Caribou/Reindeer Butchery and Consumption

Caribou butchery practices have varied widely among groups. However, common variables in the butchery process have involved transportation requirements and whether immediate or short-term consumption or storage of meat was anticipated. The size of the packages of meat produced would often be related to whether humans, dogs, or sleds (or, today, snowmobiles) were used for transportation. If carcasses were

to be used for long-term caching (storage in an underground or rock-lined chamber), they would only be beheaded, eviscerated, and quartered, with further butchery taking place before consumption at a later time. Butchery procedures may also have been affected by whether the animal was being used for purposes other than human consumption (such as clothing manufacture, dog food, or bait for traps).

Arctic peoples compensated for the relative leanness of caribou meat by the way in which the animal was butchered, cooked, and prepared. The final preparation of food sections was strongly related to preference for body parts. These were (and are) fairly constant cross-culturally and were based on the amounts of meat in different parts of the carcass, particularly on fat concentrations and the reliability of those concentrations. In general, because of higher fat content, portions nearer the axial skeleton (ribs and vertebrae), including rib steaks and roasts, were highly desired. Also high on the preference list was the backstrap (backbone musculature), followed by front (brisket, short ribs) and back (pelvis) portions. Rear haunches were generally preferred to front shanks for the same reasons. Kidneys were also prized above other viscera because of higher fat content, followed by the heart and lungs. The most desirable smaller parts included those of the head, particularly the brain, the tongue, and the fat behind the eyes. Intestinal fat was also consumed (Spiess 1979).

If alternative resources were few, a variety of techniques were employed to maximize the nutritional yield from caribou carcasses. One was the smashing or cracking of the articular ends of long bones to obtain marrow, a high-fat, high-protein substance enjoyed by hunting peoples. Afterward, the marrow and bone fragments were sometimes boiled to produce "bone juice." Data from Eskimo informants suggest that marrow was cherished as a source of fat because of high concentrations of low-melting-point fats, particularly oleic acid (Binford 1978). Bones chosen first for marrow extraction were the tibia, the "cannon bones" (metacarpals and metatarsals), the radiocubitis, the femur, and the humerus (Binford 1978: 42). Less desirable as sources of marrow were the pelvis, phalanges, mandible, scapula, and ribs. Thin animals (including nursing females or rut-depleted males) were often avoided because starvation removes fat reserves from the bone marrow.

Both marrow and caribou back-fat have been key ingredients in the traditional dish akutuk, an excellent source of vitamins as well as protein and fat. This dish, colloquially known as "Eskimo ice cream," was prepared in the past by mixing caribou tallow with berries of various types and snow. Depending on the availability of other fats, fish, or sea mammals, oils might also have been added. Today, sugar has become another ingredient in the mix.

A second major nutritional resource obtained by carcass reduction was caribou "bone grease," that is,

Table II.G.4.1. *Weights and edible weights for caribou (Rangifer tarandus)*

Age/Size Group	Weight (kg)	Percentage edible	Meat (kg)	Fat (kg)	Viscera (kg)
Fawn	11	58	3.8	1.1	2.2
Juvenile	34	57	11.8	3.4	6.8
Young adult (female)	57	56	19.8	5.7	11.4
Young adult (male)	81	56	28.1	8.1	16.2
Average adult (female)	72	50	25.5	7.3	14.5
Average adult (male)	110	50	38.2	10.9	21.8
Old adult (female)	84	39	29.2	8.4	16.8
Old adult (male)	153	39	53.1	15.3	30.6

Sources: Hill (1967), Hall (1971), Binford (1978), Spiess (1979), and Yesner (1980).

fats obtained from bone tissue itself (Leechman 1951). These were rendered by first smashing bones into small fragments, then simmering them in a pot of water and skimming off the fat. If not consumed immediately, the fat was generally stored in a cleaned caribou stomach. The best bone grease, referred to as "white" bone grease by Eskimo informants (Binford 1978), was rendered from long bones, whereas the axial skeleton, mandible, sternum, and pelvis were used to produce a less desired "yellow grease." Obtaining both marrow and bone grease could also have been part of the same procedure that employed long bones for tool manufacture (Yesner and Bonnicksen 1979). Carcass reduction for marrow and bone grease most likely took place at times of the year when caribou were scarce and/or stored resources were depleted, particularly in the late winter and early spring.

A final resource that could sometimes be obtained from caribou or reindeer were the stomach contents, that is, fermented plants of various sorts, including lichens. Although these plants cannot be readily digested by humans because of their high cellulose content, they are edible if fermented in the caribou rumen. As such, they can be an excellent source of ascorbic acid (vitamin C), otherwise obtainable only from berries (fresh, frozen, or preserved in fish or sea-mammal oils) or raw meat.

Food Storage and "Social Storage"

Many groups stored their caribou meat. Variables affecting storage included the quantity of meat to be butchered or transported and the ambient temperature at the time of the kill and afterward. There were two basic practices: the caching of unbutchered or roughly butchered meat sections and storage following butchering of the animals. Caching was likely to

occur when large numbers of caribou were killed at one time and the carcasses could not be consumed or transported immediately. This generally took place in late fall or winter, and landmarks were left behind so that the meat could be recovered even if the caches were covered by snow. Meat sections were recovered when the area was revisited during an annual round, whereupon secondary butchering would take place. Unbutchered, cached meat sections were sometimes fed to the dogs, particularly during the winter.

In situations where butchering preceded storage, the hunters employed drying and also (when possible) freezing. Specific butchery methods depended on storage techniques. Drying methods utilized drying racks (Figure II.G.4.2), whereas frozen storage required deep ice cellars, often delved out of the permafrost, that could only be used during winter or spring.

Widespread food sharing was also a technique that helped to maximize both the nutrition obtained from the caribou and the survivability of the caribou-hunting group. In groups using mass killing techniques, the individuals in charge of the hunt often acted as decision makers regarding the division of the meat. Sharing beyond the nuclear family was less likely during periods of relative abundance, but at other times of year (when individual hunting took place) food sharing was ubiquitous. It was widely recognized by various caribou-hunting groups that men differed significantly in their hunting abilities (or luck), and kills were shared with all other members of the group. In such cases, the successful hunter was often in charge of meat distribution.

Finally, trade was a mechanism by which caribou hunters obtained other foods. Among the Nunamiut of northern Alaska and the Tareumiut peoples on the North Alaskan coast, for example, long-distance trading partnerships were developed. During annual ses-

Figure II.G.4.2. A spring drying rack for caribou meat used by the Nunamiut Eskimo of the Brooks Range, northern Alaska. (From Binford 1978: 99; used by permission of Academic Press.)



sions, frequently held during summer months, caribou hides were traded for dried fish, seal meat, and fish and sea-mammal oils. Similar patterns evolved in Scandinavia and Russia.

The Domestication of Reindeer

The domestication of reindeer marked a considerable departure from previous dependence on wild reindeer (caribou). Although human manipulation of caribou populations is thousands of years old, true domestication of reindeer began no more than 400 to 500 years ago in northern Scandinavia and Siberia (Anderson 1958; Vorren and Manker 1962; Levin and Potapov 1964). O. Magnus (1555) is probably the earliest historical source referring to reindeer domestication by Saami (Lapps). However, there is some archaeological evidence from medieval Scandinavia to suggest that this transition may have taken place as early as the beginning of the fifteenth century (Simonsen 1972; Hambleton and Rowly-Conwy 1997). Historical sources suggest that reindeer domestication began with the expansion of peoples from the south into the territory of these northern peoples, including Norse populations in Scandinavia and Evenks (Tungus) in Siberia. These expansions were probably coincident with a climatic optimum in the sixteenth century. The result was removal of reindeer habitat, increased density of northern peoples, and overhunting of reindeer populations, in part resulting from the introduction of firearms. Domestication apparently occurred in order to provide northern groups with a reliable source of food in the face of the decline of reindeer populations. Unlike other northern groups, the Saami regularly milked their reindeer, producing a kind of cheese (Magnus 1555; Utsi 1948). By the seventeenth century, Saami peoples were regularly trading reindeer hides for grain, salt, metal objects, cloth, and alcohol (Schefferus 1674). In part, this trading was to pay for the taxation of reindeer herds imposed by Norse authorities.

Historical sources also indicate that initially the domestication of reindeer (and their herding and use as draft animals) was undertaken as an aid to hunting before it became a sustainable livelihood. Also, in some areas, such as parts of northern Siberia, reindeer herding remained a hunting aid and/or source of reserve food into the twentieth century. Later, more closely supervised and intensive herding took place, beginning no more than 200 years ago. Such herding involved draft reindeer for carrying loads and for riding. As with wild reindeer hunting, domesticated reindeer herding has involved the use of a wide variety of techniques for corralling and killing animals, including, as already mentioned, the use of rifles since the late nineteenth century. At the same time, like cattle herding, it has also involved the use of a wide variety of techniques for breaking animals and for indicating ownership (for example, ear notching).

The necessity for continuous human contact in herding, caring for, and driving reindeer populations created an ever increasing dependence of humans on the animals. This, in turn, necessitated greater reindeer-herding specialization, a more sedentary lifestyle, and, in some cases, a discontinuance of customary transhumant movements into coastal or riverine zones for fishing and/or sea-mammal hunting. The result was the differentiation in Scandinavia between the so-called Mountain Lapps or Reindeer Lapps (focused on reindeer herding), the Forest Lapps, and the Coast Lapps or Fisher Lapps. In Siberia, a similar process can be seen in the evolution of the "Reindeer Chukchi." Development of discrete groups of herders also took place, with differences in reindeer paraphernalia identifying such groups. Problems have also developed in mating or interaction between domesticated reindeer herds and wild reindeer populations, necessitating special efforts to separate these groups.

Reindeer Husbandry in the Twentieth Century

In northwestern Alaska, reindeer herding began in the 1920s with the noted Protestant minister Sheldon Jackson, who introduced the animals into the area (Scheffer 1951; Stern et al. 1980). Jackson believed that reindeer would help to augment the diet of the Bering Strait Inupiat (Northern Eskimo) people, some of whose other resources, especially large baleen whales, had been depleted by the ingress of other populations since the mid-nineteenth century. Jackson brought in Saami (Lapp) herders to help the Alaskan natives learn about reindeer (Vorren 1994). Although the herds depleted some habitats and caused disruption of traditional settlement patterns and lifestyles, reindeer herding continues today in that part of Alaska, where it is a successful enterprise managed by the Northwest Alaska Native Association (NANA).

In the former Soviet Union, large cooperatives, equivalent to communal farms, were established for the marketing and distribution of reindeer meat. Thus, much of the meat was not consumed by local peoples but by others in the region. Payments for reindeer meat were handled by local officials, assisted by Communist Party advisory groups. Such practices have been terminated with the fall of the Soviet Union.

In the latter part of the twentieth century, more extensive and larger-scale reindeer husbandry or ranching has been developed in a number of places, including Alaska, northwestern Canada, western Greenland, and Siberia (Strickon 1965; Sturdy 1972; Ingold 1980). This development has everywhere required the use of modern mechanized equipment, including trucks, airplanes, and particularly snowmobiles (Pelto 1973; Beach 1981; Paine 1994). There have been problems in adapting mechanized systems to reindeer herding because of the traditional skittishness of the animals. Such systems have also continued

to transform traditional settlement patterns and lifestyles. The marketing and sale of meat has been initiated in a number of locales, particularly in Scandinavia, Greenland, and Russia. The sale of reindeer meat is more restricted in Canada and Alaska; in the latter region, however, it is a key ingredient in "reindeer sausage," which is widely marketed. The meat is frequently leaner and milder in flavor than that of wild caribou and is preferred by many non-native consumers for that reason. A key reason for the present-day maintenance of reindeer herds in a number of places, including Alaska and Russia, is the strong demand for - and purchase of - reindeer antlers in East Asia. There, the antlers, alleged to possess the properties of a stimulant and an aphrodisiac, are ground to a pharmaceutical powder and retailed to the public.

Survival of a Subsistence Based on Caribou/Reindeer

Both herding and hunting of reindeer have continued to provide meat for indigenous peoples throughout the northern circumpolar region and have proven to be important local sources of human nutrition. To a lesser extent, caribou or wild reindeer are also an important resource for non-native peoples who hunt them and consume their meat. Thus, in Alaska, special permits are required for non-native caribou hunting, and preference is given to those who can demonstrate past reliance on this meat to feed their families. The use of caribou and reindeer is governed by local and national laws. In addition, international treaties, such as that between the United States and Canada, govern the management and exploitation of caribou herds that cross national boundaries. In Scandinavia, such treaties have ensured the movement of traditional Saami herds and herders across the northern part of the region.

Unfortunately, there are a number of threats to the continued use and consumption of caribou and reindeer meat by northern peoples. Habitat disruption has probably had the greatest effect on both wild and domestic reindeer. For example, even in those areas where traditional caribou hunting persists, there is a significant difference between the hunting patterns of the past and those of today. Illustrative is Alaskan archaeological data, which demonstrates that traditional peoples hunted caribou as much as 15 or more years old, but animals of this age are rarely seen in wild populations today (Yesner 1980) (see Table II.G.4.2).

Significant alterations of caribou and reindeer habitats are the result of a number of factors: the increased movement of non-native peoples into arctic regions; the development of oil, mining, and other industries in the Arctic (particularly in Alaska and Russia); and the construction of large hydroelectric projects (particularly in Canada and Scandi-

Table II.G.4.2 Age distribution of caribou from an Alaskan archaeological site

Age	Number of individuals
0-12 months (0-1 yr)	14
12-24 months (1-2 yr)	10
24-36 months (2-3 yr)	9
36-48 months (3-4 yr)	27
48-60 months (4-5 yr)	10
60-72 months (5-6 yr)	6
72-96 months (6-8 yr)	8
96-120 months (8-10 yr)	1
120-156 months (10-13 yr)	3
156-180 months (13-15 yr)	5
180-204 months (15-17 yr)	1
204+ months (17+ yr)	1
<i>N</i> = 87	

Source: Yesner (1980), p. 31.

navia). The latter has had a particularly strong impact by flooding large areas of the traditional caribou range. In addition, air pollution, particularly that resulting from the Chernobyl reactor accident in the former Soviet Union, has affected the caribou. It is too early to determine the long-term consequences for the subsistence of northern peoples, particularly in Scandinavia and Russia, but reindeer meat will probably continue to be important in these regions for some time. In Alaska and Canada, although some herds have been reduced, others still number in the hundreds of thousands. It is to be hoped that reindeer and caribou will continue to be significant dietary resources well into the twenty-first century.

David R. Yesner

Bibliography

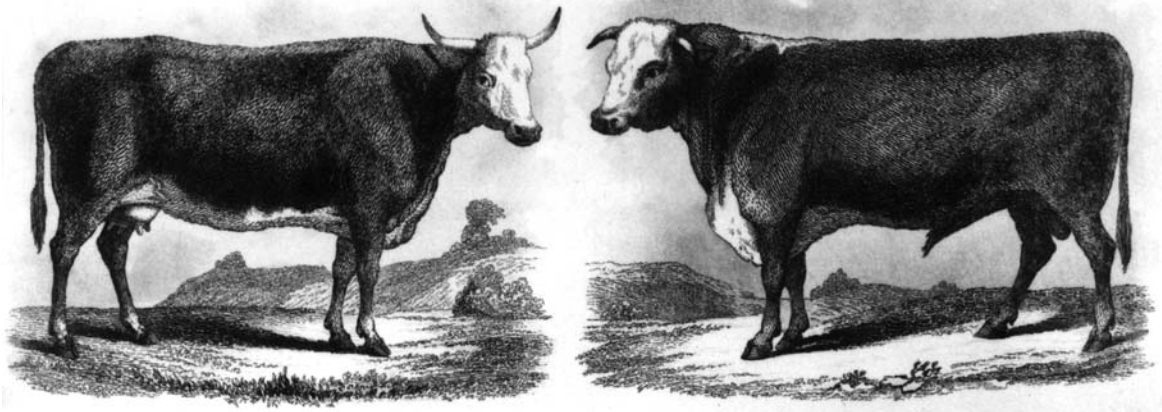
- Anderson, R. T. 1958. Dating reindeer pastoralism in Lapland. *Ethnohistory* 5: 361-91.
- Anell, Bergt. 1969. Running down and driving of game in North America. *Studia Ethnographica Upsaliensia* 30: 1-300.
- Arima, Eugene Y. 1975. *A contextual study of the Caribou Eskimo kayak*. National Museum of Man, Mercury Series No. 5, Ottawa.
- Banfield, A. W. F. 1961. *A revision of the reindeer and caribou genus Rangifer*. Ottawa.
- Beach, Hugh. 1981. *Reindeer herd management in transition*. Uppsala Studies in Cultural Anthropology No. 3. Uppsala, Sweden.
- Binford, Lewis R. 1978. *Nunamiut ethnoarchaeology*. New York.
- Birket-Smith, D. 1929. *The Caribou Eskimos*. Report of the Fifth Thule Expedition, Vol. 5, Part 1. Copenhagen.
- Blehr, Otto. 1990. Communal hunting as a prerequisite for caribou (wild reindeer) as a human resource. In

- Hunters of the recent past*, ed. Leslie B. Davis and Brian O. K. Reeves, 304–26. London.
- Bouchud, Jean. 1966. *Essai sur le renne et la climatologie du paléolithique moyen et supérieur*. Périgueux, France.
- Boyle, Katharine V. 1990. *Upper Palaeolithic faunas from Southwest France: A zoogeographic perspective*. British Archaeological Reports, International Series No. 557. Oxford.
1993. Upper Paleolithic procurement and processing strategies in Southwest France. In *Hunting and animal exploitation in the Later Paleolithic and Mesolithic of Eurasia*, ed. G. L. Peterkin, Harvey M. Bricker, and Paul Mellars, 1–10. Archaeological Papers of the American Anthropological Association No. 4. Washington, D.C.
- Buikstra, Jane E. 1976. The Caribou Eskimo: General and specific disease. *American Journal of Physical Anthropology* 45: 351–68.
- Burch, Ernest L., Jr. 1972. The caribou/wild reindeer as a human resource. *American Antiquity* 37: 339–68.
- Burke, Ariane, and Anne Pike-Tay. 1997. Reconstructing “l’Age du Renne.” In *Caribou and reindeer hunters of the Northern Hemisphere*, ed. Lawrence J. Jackson and Paul T. Thacker, 69–81. Aldershot, England.
- Chapdelaine, Claude, ed. 1985. Des éléphants, des caribous, et des hommes: La Période Paléolindienne. *Recherches Amerindiennes au Québec* 15: 1–185.
- Chard, Chester S. 1963. The Nganasan: Wild reindeer hunters of the Taimyr Peninsula. *Arctic Anthropology* 1: 105–30.
- Cleland, Charles. 1965. Barren ground caribou (*Rangifer arcticus*) from an Early Man site in southeastern Michigan. *American Antiquity* 30: 350–1.
- David, François, and J. G. Enloe. 1992. Chasse saisonnière des magdaleniens de Bassin Parisien. *Bulletins et Mémoires Société d’Anthropologie de Paris* 4: 167–74.
- David, Nicholas C. 1973. On Upper Paleolithic society, ecology, and technological change: The Noaillian case. In *The explanation of culture change*, ed. Colin Renfrew, 277–303. London.
- Delpech, Françoise. 1983. *Les faunes du Paléolithique Supérieur dans le sud-ouest de la France*. Cahiers du Quaternaire No. 6. Bordeaux.
- Delpech, Françoise, and E. Heintz. 1976. Les artiodactyles: Cervides. In *La Préhistoire Française*, Vol. 1, ed. Henri de Lumley, 395–404. Paris.
- Draper, Harold H. 1974. Aspects of adaptation to the native Eskimo diet: Energy metabolism. *American Journal of Physical Anthropology* 43: 475–85.
- Enloe, J. G., and François David. 1997. *Rangifer* herd behavior: Seasonality of hunting in the Magdalenian of the Paris basin. In *Caribou and reindeer hunters of the Northern Hemisphere*, ed. Lawrence J. Jackson and Paul T. Thacker, 52–68. Aldershot, England.
- Fitzhugh, William W. 1972. Environmental archaeology and cultural systems in Hamilton Inlet, Labrador. *Smithsonian Contributions to Anthropology* 16: 1–300.
- Foulks, Edward F. 1972. *The arctic hysterias of the North Alaskan Eskimo*. American Anthropological Association, Anthropological Studies No. 10. Washington, D.C.
- Funk, Robert E., D. W. Fisher, and E. M. Reilly, Jr. 1970. Caribou and Paleo-Indian in New York State: A presumed association. *American Journal of Science* 268: 181–6.
- Gordon, Bryan H. C. 1975. *Of men and birds in barrenland prehistory*. National Museum of Man, Mercury Series, Archaeological Papers No. 28. Ottawa.
1988. *Of men and reindeer birds in French Magdalenian prehistory*. British Archaeological Reports, International Series No. 390. Oxford.
1990. World *Rangifer* communal hunting. In *Hunters of the recent past*, ed. Leslie B. Davis and Brian O. K. Reeves, 277–303. London.
- Greenman, E. F., and George M. Stanley. 1977. Prehistoric Chipewyan harvesting at a Barrenland caribou water crossing. *Western Canadian Journal of Anthropology* 7: 69–83.
- Gronnow, Bjarne, Morten Meldgaard, and Jorn Berglund Nielsen. 1983. Aasivissuit – The great summer camp: Archaeological, ethnographic, and zooarchaeological studies of a caribou hunting site in West Greenland. *Meddelelser om Groenland, Man and Society* 5: 1–96.
- Gubser, Nicholas J. 1965. *The Nunamiut Eskimos: Hunters of caribou*. New Haven, Conn.
- Hall, Edwin S., Jr. 1971. Kangigugsuk: A cultural reconstruction of a sixteenth century Eskimo site in Northern Alaska. *Arctic Anthropology* 8: 1–101.
- ed. 1989. *People and caribou in the Northwest Territories*. Yellowknife, Canada.
- Hambleton, Ellen, and Peter Rowley-Conwy. 1997. The medieval reindeer economy in the Varanger Fjord, North Norway. *Norwegian Archaeological Review* 30: 55–70.
- Heard, Douglas C. 1997. Causes of barren-ground caribou migrations and implications to hunters. In *Caribou and reindeer hunters of the Northern Hemisphere*, ed. Lawrence J. Jackson and Paul T. Thacker, 27–31. Aldershot, England.
- Hill, Richard. 1967. *Mackenzie reindeer operations*. Ottawa.
- Ingold, Timothy. 1980. *Hunters, pastoralists, and ranchers*. Cambridge and New York.
- Irimoto, Takashi. 1981. *Chipewyan ecology: Group structure and caribou hunting system*. Senri Ethnological Studies No. 8. Osaka.
- Jackson, Lawrence J. 1988. Fossil cervids and fluted point hunters. *Ontario Archaeology* 48: 27–41.
- Kelsall, John P. 1968. *The migratory barren-ground caribou of Canada*. Canadian Wildlife Service Monograph No. 3. Ottawa.
- Krupnik, Igor. 1993. *Arctic adaptations: Native whalers and reindeer herders of northern Eurasia*. Hanover, N.H.
- Laughlin, William S. 1966. Genetic and anthropological characteristics of arctic populations. In *The biology of human adaptability*, ed. Paul T. Baker and J. S. Weiner, 469–96. Oxford.
- Leechman, D. 1951. Bone grease. *American Antiquity* 16: 355–6.
- Levin, Michael G., and L. P. Potapov, eds. 1964. *The peoples of Siberia*. Chicago.
- Lips, J. 1947. Notes on Montagnais-Naskapi economy (Lake St. John and Lake Mistassini bands). *Ethnos* 12: 1–78.
- Loring, Stephen J. 1997. On the trail to the caribou house: Some reflections on Innu caribou hunters in northern Labrador. In *Caribou and reindeer hunters of the Northern Hemisphere*, ed. Lawrence J. Jackson and Paul T. Thacker, 185–220. Aldershot, England.
- Magnus, O. 1555. *History of the northern peoples (Historia de gentibus septentrionalibus)*. Rome.
- Mann, G. V., E. M. Scott, L. M. Hursh, and C. A. Heller. 1962. The health and nutritional status of Alaskan Eskimos. *American Journal of Clinical Nutrition* 11: 31–66.
- Meldgaard, Morten. 1983. Resource fluctuations and human subsistence: A zooarchaeological and ethnographic investigation of a West Greenland caribou hunting camp. In *Animals and archaeology: Hunters and*

- their prey*, ed. Juliet Clutton-Brock and Caroline Grigson, 259–72. British Archaeological Reports, International Series No. 163. Oxford.
1986. *The Greenland caribou: Zoogeography, taxonomy, and population dynamics*. Copenhagen.
- Mellars, Paul R. 1996. *The Neanderthal legacy: An archaeological perspective from western Europe*. Princeton, N.J.
- Messier, F., J. Huot, D. LeHenaff, and S. Luttich. 1988. Demography of the George River caribou herd: Evidence of population regulation by forage exploitation and range expansion. *Arctic* 41: 279–87.
- Moore, Omar Khayyam. 1969. Divination: A new perspective. In *Environment and cultural behavior*, ed. Andrew P. Vayda, 121–9. Austin, Tex.
- Morrison, David. 1982. Chipewyan drift fences and shooting blinds in the central Barren Grounds. In *Megaliths to medicine wheels: Boulder structures in archaeology*, ed. Michael Wilson, K. L. Road, and K. J. Hardy, 171–85. Calgary.
- Mowat, Farley. 1952. *People of the deer*. London.
- Nelleman, G. 1970. Caribou hunting in West Greenland. *Folk* 12: 133–53.
- Paine, Robert. 1994. *A portrait of Saami reindeer pastoralism*. Washington, D.C.
- Parker, G. 1971. *Trends in the population of barren-ground caribou over the last two decades: A re-evaluation of the evidence*. Canadian Wildlife Service, Occasional Paper No. 10. Ottawa.
- Peers, Laura. 1986. Ontario Paleo-Indians and caribou predation. *Ontario Archaeology* 43: 31–40.
- Pelto, Pertti J. 1973. *The snowmobile revolution: Technology and social change in the Arctic*. Prospect Heights, Ill.
- Pike-Tay, Anne, and Harvey M. Bricker. 1993. Hunting in the Gravettian: An examination of evidence from south-western France. In *Hunting and animal exploitation in the Later Palaeolithic and Mesolithic of Eurasia*, ed. Gail Larsen Peterkin, Harvey M. Bricker, and Paul Mellars, 127–44. Archaeological Papers of the American Anthropological Association No. 4. Washington, D.C.
- Pruitt, W. O. 1960. Behavior of the barren-ground caribou. *Biological Papers of the University of Alaska* 3: 1–43.
- Scheffer, Victor B. 1951. The rise and fall of a reindeer herd. *Scientific Monthly* 73: 356–61.
- Schefferus, J. 1674. *The history of Lapland*. Oxford.
- Simchenko, Yri B. 1976. *The culture of reindeer hunters of northern Eurasia*. In Russian. Moscow.
- Simonsen, P. 1972. The transition from food-gathering to pastoralism in North Scandinavia and its impact on settlement patterns. In *Man, settlement, and urbanism*, ed. Peter J. Ucko, Ruth Tringham, and Geoffrey W. Dimbleby, 431–9. London.
- Skoog, Robert O. 1968. *Ecology of the caribou in Alaska*. Ann Arbor, Mich.
- Smith, James G. E. 1978. Economic uncertainty in an “original affluent society”: Caribou and caribou-eater Chipewyan adaptive strategies. *Arctic Anthropology* 15: 68–88.
- Spies, Arthur E. 1979. *Reindeer and caribou hunters: An archaeological study*. New York.
- In press. Étude de la saison d’habitation au moyen du ciment dentaire. In *Les derniers chasseurs du renne le long des Pyrénées*, ed. Lawrence Guy Straus. *Mémoires de la Société Préhistorique Française*. Paris.
- Spies, Arthur E., Mary Lou Curran, and John Grimes. 1985. Caribou (*Rangifer tarandus*) bones from New England Paleo-Indian sites. *North American Archaeologist* 6: 145–59.
- Stern, Richard O., Edward L. Arobio, Larry I. Naylor, and Wayne Thomas. 1980. Eskimos, reindeer, and land. *University of Alaska Agricultural Experiment Station Bulletin* 59: 1–205.
- Straus, Lawrence G. 1997. Late Glacial reindeer hunters along the French Pyrenees. In *Caribou and reindeer hunters of the Northern Hemisphere*, ed. Lawrence J. Jackson and Paul T. Thacker, 165–84. Aldershot, England.
- Strickon, A. 1965. The Euro-American ranching complex. In *Man, culture, and animals*, ed. Anthony Leeds and Andrew Vayda, 229–58. Washington, D.C.
- Sturdy, D. A. 1972. The exploitation patterns of a modern reindeer economy in West Greenland. In *Papers in economic prehistory*, ed. Eric S. Higgs, 161–8. Cambridge.
1975. Some reindeer economies in prehistoric Europe. In *Palaeoeconomy*, ed. Eric S. Higgs, 55–96. Cambridge and New York.
- Thacker, Paul T. 1997. The significance of *Rangifer* as a human prey species during the central European Upper Paleolithic. In *Caribou and reindeer hunters of the Northern Hemisphere*, ed. Lawrence J. Jackson and Paul T. Thacker, 82–104. Aldershot, England.
- Utsi, M. 1948. The reindeer-breeding methods of the northern Lapps. *Man* 48: 97–101.
- Vorren, Ornulov. 1994. *Saami, reindeer, and gold in Alaska*. Prospect Heights, Ill.
- Vorren, Ornulov, and Ernst Manker. 1962. *Lapp life and customs: A survey*. London.
- White, Theodore. 1953. A method of calculating the dietary percentage of various food animals utilized by aboriginal peoples. *American Antiquity* 18: 396–8.
- Yesner, David R. 1980. Caribou exploitation in interior Alaska: Paleoeconomy at Paxson Lake. *Anthropological Papers of the University of Alaska* 19: 15–31.
1989. Moose hunters of the boreal forest? A reconsideration of subsistence patterns in the western Subarctic. *Arctic* 42: 97–108.
1994. Seasonality and resource stress among hunter-gatherers: Archaeological signatures. In *Key issues in hunter-gatherer research*, ed. Ernest L. Burch and Linda J. Ellanna, 99–123. New York.
1995. Subsistence diversity, faunal extinction, and hunter-gatherer strategies in Late Pleistocene/Early Holocene Beringia: Evidence from the Broken Mammoth site, Big Delta, Alaska. *Current Research in the Pleistocene* 11: 154–6.
- Yesner, David R., and Robson L. Bonnicksen. 1979. Caribou metapodial shaft splinter technology. *Journal of Archaeological Science* 6: 303–8.

II.G.5 Cattle

Although cattle have been domesticated for less than 10,000 years, they are the world’s most important animals, as judged by their multiple contributions of draft power, meat, milk, hides, and dung. In Asia and Africa, the tie between man and beast is much more than economic (as it is in the West), and domestication itself seems to have occurred for noneconomic reasons. Cattle, like other ruminants, convert cellulose-rich materials



Cattle

– that are otherwise useless to humans as food – into carbohydrates, fats, and proteins. (In industrialized countries, however, cattle are fed grains from cultivated land and consequently can be viewed as competing for foods that could go directly to humans.)

The term “cattle” can have a broad or narrow meaning. One usage subsumes all five domesticated species in the genus *Bos* as cattle. The other restricts the term to only the two main bovines in this genus: European cattle (*Bos taurus*) and zebu cattle (*Bos indicus*). Both animals were derived from the same wild ancestor, the aurochs (*Bos primigenius*), and they can interbreed to produce fertile offspring. Three much more localized Asian species sometimes fall under the rubric of domesticated cattle: mithan (*Bos frontalis*), yak (*Bos grunniens*), and banteng (*Bos javanicus*). The mithan is found in a forested region that encompasses north-east India, northwest Burma, and Bhutan. Its ancestor is the wild gaur (*Bos gaurus*). The yak of Nepal and Tibet was derived from the wild yak (*Bos mutus*), whereas the banteng (“Bali cattle”), found on several islands of the Indonesian archipelago, was domesticated from the wild banteng of the same species. Unless otherwise qualified, the term “cattle” in this chapter refers to the two main species derived from the aurochs.

Domestication

Evidence for the domestication of cattle dates from between 8,000 and 7,000 years ago in southwestern Asia. Such dating suggests that cattle were not domesticated until cereal domestication had taken place, whereas sheep and goats entered the barnyard of humans with the beginning of agriculture. Although the aurochs has been extinct since the seventeenth century, its role as the wild ancestor of cattle has never been seriously disputed. This impressive beast

occurred throughout Eurasia south of the taiga and north of the desert and tropical forest. Paleolithic cave paintings at Lascaux and Altamira in Europe convey a sense of its fierce nobility, which must have awed hunters. At Mesolithic sites in Eurasia, the relative paucity of aurochs bones compared to those of wild sheep – a difficult quarry – and goats suggests a major differential in hunting success.

That humans with primitive technology would have even attempted to tame such a ferocious animal becomes plausible only if one constructs a lure scenario. This setting would have the creation of artificial salt licks to entice the aurochs to a place where the opportunity might present itself to steal a calf from its mother. A young animal, of course, could have been captured and tamed much more easily than an adult.

It is probable that the aurochs underwent transformation into cattle more than once in the prehistory of the Old World. Mitochondrial DNA research indicates that zebu cattle of India were domesticated independently from European cattle and from a different subspecies of the aurochs (Loftus et al. 1994). The clustering of all African zebu within the *B. taurus* lineage is based on the assumption that a zebu ancestor crossbred with earlier *B. taurus* in Africa, and the zebu cattle themselves may have had more than one place of origin. A humped bovid on a Near Eastern figurine dates from the late Bronze Age and corresponds to a similar kind of animal found at Mehrgarh in Pakistan. At that site, both zebu and European cattle are found and together comprise more than 50 percent of the mammalian assemblage (Meadows 1984). Although their origins seem to have been outside the tropics, zebu are well adapted to hot climates, perhaps because much of their early husbandry was in India. Domestication may have enhanced the zebu's characteristic hump, dewlap, generally white color, and alert nature.

Early Neolithic cattle keeping in northern Africa points to a possible separate domestication there as well (Close and Wendorf 1992), and S. Bökönyi (1976) has gone so far as to assert that domestication of cattle from the wild happened repeatedly as late as the classical period of history. Indeed, the first-century-B.C. Roman poet Virgil stated that after a virulent disease had decimated the herds, peasants transformed the aurochs into a domesticated bovine to replace their previous stock. Discrete domestications may have occurred in several places, but the diffusion of domesticated cattle would have been a much more important overall process than repeatedly transforming a ferocious creature over many generations.

As with sheep and goats, the process of domesticating cattle resulted in animals smaller than the wild progenitor. Dated osteological material from Neolithic sites establishes the transition from wild to domesticated, although the sexual dimorphism between the male and female aurochs was not originally appreciated, and the "wild" label was misapplied to the former and "domesticated" to the latter (Grigson 1969). The Fertile Crescent has long been considered the place of initial cattle domestication, but that view tends to reflect the large number of excavations made there. Early signs of Neolithic cattle keeping have also been found in Anatolia (Turkey), where the osteological material at Catal Hüyük provides evidence of the transition from the aurochs of 8,400 years ago to cattle by 7,800 years ago. In short, it is still premature to specify where the first cattle were domesticated.

The extraordinary usefulness of cattle would superficially seem to have been the motivation for their domestication. In other words, given all the benefits that cattle impart, it was logical that the aurochs would come under human control, which is an extension of a deeply rooted Western concept that nature exists to serve the practical needs of people and that necessity has always elicited human ingenuity to provide technical solutions.

However, an alternative to such a materialistic perspective was first proposed by Eduard Hahn (1896), a German geographer-ethnologist. Hahn probed beneath the shallow surface of materialistic motives to see noneconomic forces at work in transforming the animal into what is essentially an artifact of artificial selection. He noted that the curved horns of the aurochs resembled the lunar crescent, which symbolized the Moon Goddess, and suggested that the original reason for domestication was to sacrifice these animals as a reenactment of the goddess's death, with such sacrifices perhaps performed at each waning of the moon.

Such a practice would have required a supply of animals that was initially met by capturing them from the wild. But in the holding pens, some captive bulls and cows (both with long horns) bred, and from these matings, calves occasionally were born that had physical characteristics different from their parents.

Their overall size was smaller, their temperament more docile, and their markings and hide color had unusual variations. Viewed as special, these aurochs born in captivity were also kept as objects of sacrifice but were allowed to breed, and phenotypic distinctiveness enhanced their sacred status.

Some of the next generation to follow may have reinforced the characteristics of the parents, and a gene pool that distinguished these bovines from their wild forebears gradually formed. No longer were they aurochs, but rather cattle, whose sacred roles included pulling ceremonial wagons in religious processions and symbolically plowing the land in fertility rites. Their milk was perceived to be a ritual gift from the goddess, and the most docile cows let themselves be milked by a priest in the presence of their calves. Gradually, the economic usefulness of cattle asserted itself for plowing fields, pulling wagons, and for the meat, milk, and hides, which came to be highly prized. But these were later and secondary benefits of bovine manipulation that initially was prompted by the early mythology of Near Eastern religion. The Eurasian religion of the Upper Paleolithic was supplanted by goddess cults that evolved with the emergence of agriculture and livestock husbandry into fertility cults meant to ensure the prosperity of crops and herds (Isaac 1970).

Evidence for the prehistoric importance of cattle in ritual and the noneconomic motives for domestication comes from Catal Hüyük, a Neolithic archaeological site 50 kilometers south of Konya, Turkey (Mellaart 1967). Many shrines at this reconstructed site, including the earliest one found, have bulls' heads, and an abundance of cattle horn displays suggests that they were venerated as fertility symbols. In another kind of representation, a goddess is shown as having given birth to a bull. Moving to modern times, the previously mentioned mithan (or "Indian bison," as it is sometimes called) is a domesticated animal, but it is not used for draft, meat, or milk. Rather, it is kept as an object of ritual and sacrifice, providing a contemporary example of domestication for noneconomic reasons (Simoons and Simoons 1968). Mithans – periodically killed in ritual sacrifices – occur across a continuum varying from very tame to untamed forest-dwelling animals, and salt, for which they have an insatiable craving, is the enticement that humans employ to attract them.

Early Civilizations

In Mesopotamia, Egypt, and the Indus Valley, cattle served humans as objects of ritual and sources of meat, milk, and draft power, and in all three civilizations, they were the objects of much more intensive human attention than were sheep or goats. Draft animals were indispensable for working irrigated plots in desert and semiarid regions. Cattle required water and grass and needed considerable care. Using cattle for

meat seems to have been less important than for draft. During the third Ur dynasty (2112–2004 B.C.), cattle products supplied only 10 percent of total meat needs (Adams 1981). In Mesopotamia, state authority was involved in the organization of cattle keeping, and cattle were called by different terms depending on their function (Zeder 1991).

Ritual use of cattle has been a Eurasian continuity. In ancient Egypt, cattle were regarded as sacred by several cults, particularly that of the goddess Isis, and in Crete, friezes more than 4,000 years old display evidence of bovine importance. At the palace of Minos, for example, the royal couple dressed up as a bull and a cow, and the Minoans also watched specially trained athletes who jousting bulls with acrobatic daring. The Phoenician cult of Baal had a god of fertility represented as a bull – a cult which spread to their colonies as far west as present-day Spain. The early Hebrews worshiped the bull, and in Canaan, the Baal cult was fused with the worship of Yahweh. During King Solomon's reign, the temple housed bulls of bronze and horned sculptures.

In classical antiquity, there is a rich history of bulls offered in ritual sacrifice. Some of these rituals in Greece provided a mechanism to reaffirm the social hierarchy of the polis. Sacred cattle were kept at Eleusis where Demeter was worshiped. In ancient Rome, a white bull was sacrificed at the annual *Feriae Latinae*, and its meat was distributed in proportion to the relative importance of the member cities of the Latin League. In Europe and the Middle East, Judaism, Christianity, and Islam all suppressed such sacred cattle cults as pagan manifestations that competed with their messages, although a lingering remnant is the contemporary Spanish bullfight, which reenacts a sacrifice to symbolize the eternal struggle between humans and nature. In its evolution, bullfighting lost its obvious religious content, which may explain its survival through 2,000 years of Christianity.

Europe

The westward diffusion of cattle throughout Europe was tied to the invention of the wooden plow. The harnessing of a powerful animal to that device made it possible to greatly extend cultivation without a corresponding increase in human population. At the same time, in any one community it freed surplus labor for other activities, which enabled the group to develop greater socioeconomic complexity. Castration of the bull to create an ox was a key element in the spread of the plow. Steady and strong – but docile – the ox made enormous contributions to agriculture. Osteometric evidence shows that oxen were already in use during the fourth millennium B.C., and a figure from Nemea in Greece shows yoked oxen dating from the third millennium B.C. (Pullen 1992).

Farther north in Europe, where wet summers provided abundant forage, cattle had a bigger role to play

in livestock husbandry. Following the Middle Ages, an appetite for beef grew, and, in fact, the Europeans' frenetic quest for spices was related to their growing carnivorous tastes. Until the late eighteenth century, cattle were the most important livestock in the uplands of the British Isles, where they grazed mainly on commonly owned natural pastures and secondarily were fed hay cut and stored for winter use. Other animals – sheep, pigs, goats, and horses – were made a part of a land-use system that also favored dairying and the cultivation of grains.

The relative isolation of each region resulted in locally limited gene pools for *B. taurus*, which led to different cattle phenotypes. Three of these, Aberdeen Angus, Shorthorn, and Hereford, have diffused overseas to become modern ranching stock in the Americas; other breeds, such as Devon, Skye, Galloway, Kerry, and Durham, are now rare. Characteristic of British livestock tradition was the close management and selective breeding that imparted a generally docile behavior to the animal. Fat cattle were the aesthetic ideal but also a practical outcome of the high value placed on tallow. Emphasis on pure bloodlines reflected, in part, competition among the landed gentry to produce the best animals, and, in fact, purebred cattle became a symbol of the British ruling class that dominated the world in the nineteenth century. Moreover, that class extended its insistence on purity to its own members as a way of separating the rulers from the ruled in the vast British Empire.

Africa

African bovines kept by humans are usually presumed to have originated as domesticated animals in southwestern Asia, with cattle funneled into Africa through the Sinai Peninsula and, possibly, the Straits of Bab-el Mandeb across the Red Sea, eventually spreading over vast areas of the continent. There is, however, an alternative explanation anchored in the possibility of an independent domestication of cattle in Africa. Evidence that this may have been the case has been found in the northern Sahara, in Algeria, where cave paintings dating from 5,000 to 6,000 years ago indicate a pastoral way of life based on domesticated cattle (Simoons 1971). More recently, an argument has been made for independent cattle domestication in the eastern Sahara as well (Close and Wendorf 1992). Bovine bones found in faunal assemblages have been interpreted to be the remains of domesticated stock kept for milk and blood. Deep wells, dated from the ninth millennium B.C., provided the water without which these cattle could not have survived.

In Africa, more than on any other continent, cattle raising follows a paleotechnic, precapitalistic mode that depends on the natural vegetation of common lands. The transcendental importance of cattle is most apparent in the sub-Saharan eastern half of the continent, where more than 80 million head may be found

in an arc across Sudan, Ethiopia, Uganda, Kenya, and Tanzania. Numbers would be much higher than that if nagana disease (spread by tsetse flies) did not make large areas unsuitable for cattle. The ethnographic significance of this geographic concentration was realized only after Melville Herskovits (1926) published his landmark study on the "cattle complex" in East Africa. Subsequent research among different tribal groups – the Karamajong, Nandi, Dodot, Masai, Pakot, and Turkana, among others – validated his contentions of the centrality of cattle in East African life.

Some groups are strongly pastoralist, others may also engage in agriculture. In either case, cattle, whose prestige value surpasses their economic contribution, dominate the pastoral life as well as the social and spiritual activities of each group. Every animal is named and categorized: Among the Masai, for example, each individual animal can be classified according to its matriarchal bovine lineage, a spatial pattern organized by households, and a color/age/sex physical description. A major reason for the naming of cattle in East Africa is that it manifests the affection felt toward bovines as members of the family.

Native religion in East Africa involves cattle sacrifice. Bruce Lincoln (1981) has explained this practice in terms of a "cattle cycle," in which a celestial deity gave cattle to his people. When these animals were stolen by an enemy group (less common today), warriors were enlisted to recover the stolen cattle. The cycle was completed when priests subsequently sacrificed some of the cattle in order to propitiate the celestial deity.

A textbook example of how central cattle can be to the material and social existence of a people can be seen in the Nuer people of the Sudan (Cranstone 1969). Their language is rich in terms that describe and categorize cattle by horn shape, hide color, and age. Milk is a staple food, along with millet, and cattle blood is consumed. Rawhide provides tongs, drums, shields, and bedding, and cattle bones become scrapers, pounders, and beaters. The horns are used to make spoons and spearheads. Cattle dung can be a fuel, but it is also employed in plastering walls, dressing wounds, and, when dried, as a tooth powder. Cow urine is used in washing, cheese making, and dressing skins. The meat and fat of cattle are eaten, but the ritual involved in the slaughter is as important as the food.

Fewer cattle are kept in West Africa than in East Africa, in part because it embraces more desert and forest. The relative importance of cattle varies from tribe to tribe. For example, a dearth of grass and water in the southern Sahara encourages the Tuareg to herd many more goats than cattle, but farther south in the Sahel zone, the Fulani (or Peul) people have become cattle specialists with a spiritual link to their bovines. Each nuclear family owns a herd and knows each animal by name. The Fulani have solved the problem of forage through a symbiosis with their agricultural

neighbors to the south. In the dry season, they move their herds south to graze on the stubble of harvested fields. In recompense, the cattle leave manure as fertilizer. Like the tribal peoples of East Africa, the Fulani measure prestige in terms of numbers of animals. They use common pastures, but overstocking has led to range deterioration. One solution for this problem would be the introduction of a commercial system to regularly market young cattle.

India

India has about 200 million head of cattle, more than any other country. Cattle are of great importance as draft animals and as a source of dairy products (milk, curds, ghee), which constitute a significant element of the Indian diet. Cow dung has multiple uses as fuel, fertilizer, and as a ritualistic medium; even cow urine has its sacred applications. The Hindu religion bans the exploitation of cattle for meat and for hides, and to deliberately kill a cow is considered a serious offense in a country whose politicians are committed to protecting the animals. Veneration of cattle is also reflected in the institutions elaborated to protect old cows (Lodrick 1981). At the same time, compliance has its exceptions. Certain tribal groups and some lower-caste Hindus do slaughter cattle and eat their meat.

The cow has elicited much controversy in Hinduism about its origins, for its elevation to sacred status came only after Hinduism was established. One view is that the sacredness of cattle reflects the concept of the sanctity of life (*ahimsa*) that reached Hinduism through Buddhism (Simoons 1994). Opposing this idealistic perspective is the materialist argument that cow protection was ultimately imposed to assure an ongoing pool of plow animals, without which Indian agriculture could not effectively function (Harris 1966).

The New World

Latin America

Cattle began reaching the Western Hemisphere with the second voyage of Christopher Columbus in 1494, when a few head were landed on the island of Hispaniola. They multiplied rapidly – at the same time becoming feral – and eventually were more hunted than herded. Pirates, for example, organized roundups when they wanted fresh meat. From Hispaniola, cattle were taken to other islands and to the mainlands of Central, North, and South America.

In Latin America, cattle raising became an important use of land, especially in the sparsely populated *cerrado* of Brazil, the pampas of Argentina, and the llanos of Venezuela and Colombia. The business was organized in an extensive manner, with fenceless expanses, semiferal animals, mounted cowboys, and low productivity – a system that still prevails over

large areas from Mexico to northern Argentina, although the zebu now tends to be more numerous than European cattle. Since the 1970s, cattle ranching has been responsible for most tropical deforestation, especially in Amazonia and Central America, where aggressively spreading grasses of African origin, among them *Panicum*, *Hyparrhenta*, and *Pennisetum*, have become the forage for millions of cattle. This herbaceous invasion has been termed the "Africanization of the New World tropics," and the expansion of ranching into these areas has been called the "scourge of cows" (Parsons 1970, 1988). Moreover, cattle raising as a "complex" of economic, nutritional, and social factors, and the effects of this complex on the environment and human health, has been denounced as "a new kind of malevolent force in the world" (Rifkin 1992). But from the point of view of their owners, cattle are "walking bank accounts," which become fat on grasses that grow without planting or care.

After the middle of the nineteenth century, cattle ranching moved toward an intensive British style of management on the pampas of Argentina and Uruguay. Throughout the colonial period and beyond, cattle had been raised to make dried and salted beef (*tasajo*) and for hides and tallow. But the nineteenth-century invention of evaporating beef broth to make bouillon cubes provided a new export product to stimulate cattle production, and around 1860, the pampas began to be transformed. The arrival of the Shorthorn, Aberdeen Angus, and Hereford breeds; the introduction of barbed wire fencing; the cultivation of alfalfa; the use of colian windmills to provide pumped water; and the invention and installation of refrigerated chambers on ships – all set the conditions for the export of high-quality chilled meat to Europe. Since the 1880s, beef exports have underwritten the economies of Uruguay and Argentina.

North America

North American cattle raising has its roots in two very different ranching traditions (Jordan 1993). One came from Andalusia in Spain, specifically from the Guadalquivir marshes, where bovines grazed year-round and were tended by mounted cowboys who used lassos. Transferred from Spain to the Caribbean Islands, this style of ranching was taken to the Carolinas, to the Louisiana coastal plain, and to the Mexican east coast. From Mexico, it diffused northward to southern Texas, with longhorn cattle belonging to the same breed still raised along the Guadalquivir today. Elsewhere in North America, where cold winters intervened, pastoral practices were influenced by those of the British Isles. Thus, the midwestern prairies had a cattle-raising complex that included British breeds of cattle and the use of haying, fencing, barns, and intensive herd management to create docile animals.

Most cattle in North America are "finished off" on

grain in feed lots; indeed, more than 70 percent of the grain produced in the United States is fed to cattle and other livestock. A slaughtered steer yields about 60 percent of its weight in beef products, some of which is used as pet food, but most is for human consumption. The remaining 40 percent consists of fat, bones, viscera, and hide that go into a variety of industrial and household products. Cattle tend to be much leaner today than they were a century ago. Demand for low-cholesterol meat, and the use of petroleum-based detergents rather than soaps made from tallow, have greatly reduced the market for both fatty beef and beef fat.

A growing application of animal-raising technologies can be expected to increase the efficiency of cattle raising in the United States and other industrialized countries. Artificial insemination was a common practice by the 1950s, and more recent reproductive techniques include embryo transfer, estrous cycle regulation, embryo splitting, in vitro fertilization, sperm sexing, and cloning. Other techniques promote growth: Anabolic steroids have been used on beef cattle since 1954, and fully 90 percent of feedlot cattle in the United States receive these drugs. In the European Union, steroids are still banned, which has complicated the export of American beef into that large market. For dairy cattle, bovine somatotropin, a protein hormone, increases milk production by about 12 percent. Other technologies relate to processing and marketing. Irradiation of meat kills microorganisms and extends shelf life, and the extraction of water from milk to reduce transportation costs will greatly increase the milkshed range of cities.

Consumption of Bovine Flesh

Except in the few cultures that shun beef as food, bovine flesh has often been regarded as the ultimate measure of a good diet. As Europe regained its prosperity in the three decades after World War II, per capita consumption of beef increased several times. Now, more than 40 percent of all the meat produced in Western Europe is beef. Consumers in some European countries – especially Italy, France, and Austria – are exceptionally fond of veal, which is very young bovine flesh.

In Europe, veal typically comes from calves less than 6 months old that are fed on whole milk or on a formula that includes milk products. Close to half of the bovines slaughtered in France are calves, which yield veal that normally retails for 20 percent more than beef. Veal, tender and subtly flavorful, lends itself to imaginative sauces or accompaniments: Scallopini, osso buco, saltimbocca, and Wiener schnitzel are all well-known veal dishes. In Europe, calves also provide much of the most desirable organ meats: liver, sweetbreads, kidneys, brains, and tongue. In both France and Italy, veal consumption is about one-third that of beef – four times more veal, on a per capita basis, than

is eaten in the United States, where beef has been the preferred meat since the nineteenth century.

Today, more than 40 percent of beef consumed in the United States is ground, mostly for hamburgers. The hamburger has become an icon of American popular culture, even though it is said to have originated in Hamburg, Germany. "Hamburger steak" was on Delmonico's menu as early as 1836, and there are numerous claimants for the invention of the hamburger sandwich, including a German immigrant said to have brought the idea to the United States in 1892. Certainly, it was the case that the hamburger began to be nationally known at the 1904 St. Louis World's Fair, where the sandwiches were sold by German immigrants (Trager 1995).

The hamburger – as a snack or a meal – fits into the American quest for efficiency. It cooks thoroughly in less than 8 minutes, and today, fast-food operators normally cook it even before the customer places an order. Hamburger restaurants appeared in American cities beginning in the 1920s, but their organization to maximize efficiency dates from four decades later. Giant franchise corporations with headquarters in the United States have now spread over a major part of the industrialized world. At an international level, these restaurants communicate American values of efficiency, service, and cleanliness, but for many, they also define the failure of the American culinary imagination.

The American middle class is also fond of steak, and well informed about the best cuts. Steak is a favorite restaurant choice, but it is frequently cooked at home as well, often on an outdoor grill. (In 1995, 77 percent of American households owned at least one such grill.) The United States Department of Agriculture (USDA) grading system for beef has its greatest application in steak. "Prime," with the most fat marbling, is the most flavorful, tender, and costly. Only 2 percent of all the beef produced in the country is of prime grade. "Choice" is the most widely sold grade, whereas "select" is leaner, cheaper, and not as tender. Several other countries in the world – notably Argentina – have a much higher per capita consumption of beef than the United States, but this statistic probably includes considerable amounts of organ meats. By contrast, bovine organ meat attracts few consumers in the United States.

In Japan, beef consumption is a relatively recent phenomenon. Meat eating was prohibited until shortly after the Meiji Restoration in 1882. Buddhist beliefs and Shinto influences had banned the killing and eating of four-legged animals, but fish and fowl were consumed, and the Japanese still obtain most of their protein from fish and soybeans. Sukiyaki – thinly sliced, highly marbled sirloin in soy sauce, accompanied by vegetables – has been the preferred beef dish. In the preparation of sukiyaki, the sight of blood, unacceptable in Japanese culture, was avoided. But after World War II, American influences and increasing

prosperity introduced hamburger, beef chunks in curry sauce, and skewered beef, and most recently, steak has become accepted in Japan.

Daniel W. Gade

Bibliography

- Adams, R. Mc. 1981. *Heartland of cities*. Chicago.
- Bökönyi, S. 1976. Development of early stock rearing in the Near East. *Nature* 264: 19–23.
- Close, Angela E., and Fred Wendorf. 1992. The beginnings of food production in the eastern Sahara. In *Transitions to agriculture in prehistory*, ed. Anne Birgitte Gebauer and T. Douglas Price, 63–72. Madison, Wis.
- Cranstone, B. A. L. 1969. Animal husbandry: The evidence from ethnography. In *The domestication and exploitation of plants and animals*, ed. Peter J. Ucko and G. W. Dimbleby, 247–63. Chicago.
- Grigson, Caroline. 1969. The uses and limitations of differences in the absolute size in the distinction between the bones of aurochs (*Bos primigenius*) and domestic cattle (*Bos taurus*). In *The domestication and exploitation of plants and animals*, ed. Peter J. Ucko and G. W. Dimbleby, 277–94. Chicago.
- Hahn, Eduard. 1896. *Die Haustiere und ihre Beziehungen zur Wirtschaft des Menschen*. Leipzig.
- Harris, Marvin. 1966. The cultural ecology of India's sacred cattle. *Current Anthropology* 7: 51–66.
- Herskovits, Melville J. 1926. The cattle complex in East Africa. *American Anthropologist* 28: 230–72.
- Isaac, Erich. 1970. *Geography of domestication*. Englewood Cliffs, N.J.
- Jordan, Terry G. 1993. *North American cattle-ranching frontiers*. Albuquerque, N.Mex.
- Lincoln, Bruce. 1981. *Priests, warriors and cattle: A study in the ecology of religions*. Berkeley, Calif.
- Lodrick, Deryck O. 1981. *Sacred cows, sacred places: Origins and survival of animal homes in India*. Berkeley, Calif.
- Loftus, Ronan T., D. E. MacHugh, D. G. Bradley, et al. 1994. Evidence for two independent domestications of cattle. *Proceedings of the National Academy of Sciences* 91: 2757–61.
- Meadows, Richard H. 1984. Animal domestication in the Middle East: A view from the eastern margin. In *Animals and archaeology*, Vol. 3, *Early herders and their flocks*, ed. Juliet Clutton-Brock and Caroline Grigson, 309–37. Oxford.
- Mellaart, James. 1967. *Catal Hüyük: A Neolithic town in Anatolia*. New York.
- Parsons, James J. 1970. The "Africanization" of the New World tropical grasslands. *Tübinger Geographische Studien* 34: 141–53.
1988. The scourge of cows. *Whole Earth Review* 58: 40–7.
- Pullen, D. J. 1992. Ox and plow in the early Bronze Age Aegean. *American Journal of Archaeology* 96: 45–54.
- Rifkin, Jeremy. 1992. *Beyond beef: The rise and fall of the cattle culture*. New York.
- Simoons, Frederick J. 1971. The antiquity of dairying in Asia and Africa. *Geographical Review* 61: 431–9.
1994. *Eat not this flesh: Food avoidances from prehistory to the present*. Madison, Wis.

- Simoons, Frederick J., and Elizabeth S. Simoons. 1968. *A ceremonial ox of India: The mithan in nature, culture, and history*. Madison, Wis.
- Trager, James. 1995. *The food chronology*. New York.
- Zeder, Melinda A. 1991. *Feeding cities: Specialized animal economy in the ancient Near East*. Washington, D.C.

II.G.6 🐔 Chickens

Origins

The chicken (*Gallus gallus* or *Gallus domesticus*) is generally considered to have evolved from the jungle fowl (*G. gallus*), which ranges throughout the area between eastern India and Java. Within the nomenclature, *G. domesticus* is normally used by scholars who believe in a polyphyletic origin for the domestic chicken (from *G. gallus*, *Gallus sonnerati*, and *Gallus lafayettei*), whereas *G. gallus* is used by those who support a unique origin from the various subspecies of wild *G. gallus*. Debates regarding the origin and spread of the domestic chicken focus both on its genetic basis and the “hearth area” of its initial domestication.

The osteological identification of domestic chickens has been made both on a contextual basis (i.e., the occurrence of *Gallus* bones outside of the birds’ normal wild range) and on osteometric grounds (i.e., the occurrence of bones that are larger than those of modern wild jungle fowl and therefore would seem to be the result of selective breeding). Recent research of this nature has resulted in a radical revision of the standard view of the domestication of the chicken. The presence of domestic fowl bones in third-millennium-B.C. archaeological excavation contexts at Harappa and Mohenjo-Daro in Pakistan led earlier writers (Zeuner 1963; Crawford 1984)

to assume that the chicken was first domesticated in this area. However, in 1988, B. West and B.-X. Zhou presented archaeological data showing domestic chickens to be present at China’s Yangshao and Peiligang Neolithic sites, which dated from circa 6000 to 4000 B.C. As a consequence, because wild forms of *Gallus* are entirely absent in China, and as the climate would have been inimical to them in the early Holocene, it seems likely that chickens were domesticated elsewhere at an even earlier date. In the absence of evidence from India, Southeast Asia (i.e., Thailand) has been put forward as a likely hearth area (West and Zhou 1988).

Such a hypothesis seems to square nicely with some recent research on the genetic origins of chickens. In the past, a polyphyletic view often predominated because of the phenotypic diversity of modern chickens (Crawford 1984). But quite recently, A. Fumihito and colleagues (1994) have shown convincingly that all modern chicken genes can be derived from the subspecies of *Gallus* found in northeast Thailand. This demonstration provides a geographical origin for the chicken that harmonizes with the archaeological data and, at least for the moment, disposes of the argument for polygenesis.

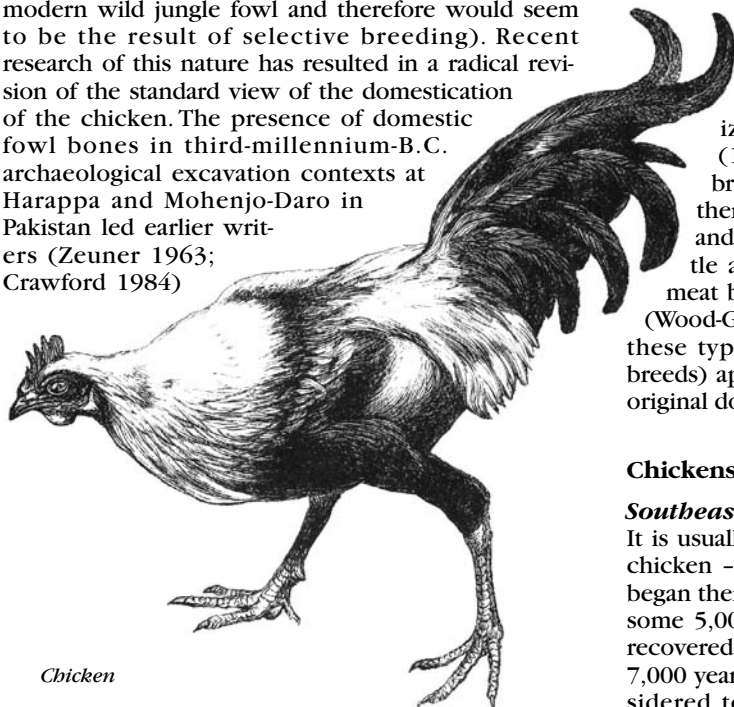
Uses

Although chickens are strongly associated with egg production in European and neo-European cultures, elsewhere they have very different associations. In much of Southeast and East Asia they have been bred both for fighting and as decoration; in Japan, for example, there is little evidence for the exploitation of chickens as food until the nineteenth century. On the basis of the present locations of these specialized breeds, I. G. Moiseeva and colleagues (1996) have postulated three centers of breed origin, to which we would add two further categories for highly productive egg layers and meat-egg compromise breeds. Both Aristotle and Pliny referred to distinct fighting and meat breeds at the beginning of the Christian era (Wood-Gush 1959). However, the origin of most of these types (save the intermediate and egg-laying breeds) appears to have been near the zone of their original domestication.

Chickens and Geography

Southeast Asia and Oceania

It is usually conceded that the Austronesians had the chicken – along with the dog and pig – when they began their epic voyages of colonization in the Pacific some 5,000 or more years ago. Chickens have been recovered from Hemudu in Southeast China, dating to 7,000 years ago, and such settlements are usually considered to be ancestral to the main Austronesian



Chicken

expansion (Bellwood 1995). Although direct evidence is lacking for chickens in the Austronesian area until much later, linguistic evidence argues that the fowl traveled with the Austronesians (Zorc 1994).

Central Asia

The chicken has been well documented osteologically from the Harappan civilization of the Indus Valley, where it was introduced by 2500 B.C. (Zeuner 1963). However, in view of both archaeological and linguistic evidence, it appears that chickens did not spread through India but rather around it – heading northeastward from China and through central Asia north of the Himalayas (Nishi 1990; Moiseeva et al. 1996). The fowl was certainly introduced into Iran about this period; Zoroastrian literature makes extensive reference to its crowing at dawn (Wood-Gush 1959).

Linguistics can complement the results of archaeology in tracking the route of the chicken's diffusion. If words for chicken, cock, and chick are compiled for the Old World, some intriguing patterns emerge. There are two extremely widespread roots, ka(C)i and tax(V).¹ The latter is spread from Korea across central Asia to the Near East, North Africa, and south to Lake Chad. This suggests not only that the chicken diffused westward from China as far as central Africa, but that it did so after the principal language phyla were established, as the vernacular terms form a chain of loanwords.

Europe

The discovery of bones – collated by West and Zhou (1988) – in central Asia seems to indicate that the chicken had reached the borders of Europe by 3000 B.C. The earliest finds come from Romania, Turkey, and Greece, where there are at least eight late Neolithic and early Bronze Age sites from which bones have been dated to the third millennium B.C. These finds, contemporary with or slightly earlier than those of the Indus Valley, appear to demonstrate that the initial dispersion of the chicken from Southeast Asia effectively bypassed the Indian subcontinent. This conclusion would seem to be further borne out by “culturally dated” fourth-millennium-B.C. finds in the vicinity of Kiev (Ukraine) that were published in the 1950s, although a great deal of research remains to be done in the area.

The subsequent rate of dispersion of the chicken in Europe appears to have slowed substantially, with most early western and northern European bone discoveries dated to the middle to late first millennium B.C. Indeed, F. Hernandez-Carrasquilla (1992) has sounded a note of caution by refuting all pre-first-millennium-B.C. identifications from Spain as intrusions or errors in attribution and has once again credited the Phoenicians with the domestic fowl's introduction into Iberia.

Chicken domestication became well established in

Europe during the Iron Age, and by the time of the Romans, superior breeding or animal husbandry strategies had resulted in substantial size increases in the domestic fowl (Thesing 1977). During the medieval period, the dietary importance of chickens increased relative to that of mammals, with chicken rearing in towns becoming increasingly common (Astill and Grant 1988; Benecke 1993). Unfortunately, chickens were often left to forage amid domestic waste, and less labor-intensive poultry-rearing techniques resulted in a general size reduction for domestic fowl (Thesing 1977), although birds bred in rural areas may have fared better. In Europe, there is little evidence for a great diversity of breeds until the late Middle Ages, or even later; in parts of the Netherlands and Poland, however, there is osteological evidence for both a small and a large breed of chicken in early medieval times (Benecke 1993).

Southwestern Asia and North Africa

In the Levant, there are only a few early chicken identifications, with the earliest finds, from Tell Swayhat in northern Syria, dating to the late third millennium B.C. (Buitenhuis 1983). The bird does not appear to have been common further south until the first millennium B.C. (West and Zhou 1988). A seal found at Nimrud, dated to 800 B.C., shows a cock (Zeuner 1963). One of the uses of early chickens in this region is indicated by the seventh-century-B.C. representation of fighting cocks on seals and potsherds from Israel/Palestine (Taran 1975).

A much-reproduced painted limestone fragment from the tomb of Tutankhamen clearly illustrates a cock, and several other representations suggest the occasional presence of fowl as exotics in Egypt during the New Kingdom (1425–1123 B.C.) (Carter 1923; Darby, Ghalioungui, and Grivetti 1977, 1: 297 ff.; Crawford 1984). However, they then disappear from the graphic record until about 650 B.C., after which they are represented in abundance (Coltherd 1966). Osteologically, despite several early misattributions, the earliest chicken bones found in Egypt date only to the beginning of the Greco-Roman period (332 B.C. to A.D. 200) (MacDonald and Edwards 1993). It would thus appear likely that the chicken was at first imported into Egypt as an exotic menagerie and fighting bird and gained economic importance only during the late first millennium B.C. – in Ptolemaic times.

Sub-Saharan Africa

Although chickens are central to African culture throughout the continent, the morphology of local fowls remains undescribed, and the routes by which they entered the continent – as well as the dates when that took place – remain poorly known. Archaeologists have tended to favor a rather recent date (MacDonald 1992, 1995; MacDonald and Edwards 1993), whereas linguists have argued for much earlier dates, based on the degree of the embedding of

words for chicken (Johnston 1886; Manessy 1972; Blench 1995; Williamson in press).

Osteological evidence of chickens in Africa has become increasingly abundant during the 1990s. The earliest known remains are from the middle of the first millennium A.D., with finds from Mali (MacDonald 1992), Nubia (MacDonald and Edwards 1993), the East African coast (Mudida and Horton 1996), and South Africa (Plug 1996) all dating to this period. No earlier remains are known as yet, and archaeozoologists hypothesize that chickens most likely entered Africa either via the Nile Valley or through an early Greco-Roman east-coast trade between about A.D. 100 and 500. A trans-Saharan introduction (at an earlier date via Phoenician Carthage), however, cannot be excluded.

The linguistic evidence rather strongly suggests multiple introductions: across the Sahara, via the Berbers, on the east coast, and, possibly, a separate introduction to Ethiopia via the Red Sea coast. In addition, early reports of chickens in Mozambique suggest black-feathered types, resembling those in India.

Because, in most cases, African chickens were left to find their own food and were allowed to mate at random, there are no significant African breeds (Kuit, Traore, and Wilson 1986; MacDonald 1992).

The New World

One of the most intriguing controversies in the history of chickens is the question of whether they were present in the New World in pre-European times. Both G. F. Carter (1971) and R. Langdon (1990) have argued strongly that they were, at least in parts of the West Coast. Linguistics (in the sense that chickens do not have names borrowed from European languages), morphology (the distinctive blue eggs of some New World breeds, otherwise known only from China), and the improbable speed of transmission inland that would be required by the assumption of a European introduction all suggest a pre-Columbian introduction from Asia (see also Carpenter 1985). Against this theory, however, is the fact that no undisputed early chicken bones have ever been found in a mainland site. Langdon (1990) has cited reports of blue-egg fowls on Easter Island, which he considers to strengthen evidence of trans-Pacific contact with the South American mainland.

Future Understanding

We know very little about past developments in five geographical areas that are crucial to understanding the domestication and subsequent spread of the domestic fowl: Thailand, Russia/Ukraine, the Indian subcontinent, Southwest Asia, and sub-Saharan Africa. It is crucial that specialized archaeozoologists examine temporally relevant avian skeletal materials from these zones and that they be well apprised of potentially confusing local wild taxa – especially birds such

as guinea fowl, pheasants, and larger francolins – and the criteria for their differentiation (MacDonald 1992).

Complementary results can be obtained from more detailed genetic work on traditional chicken breeds using both DNA and phenotypic characters. Although the chicken is usually considered to be well known, compilations such as those by Ronan Loftus and Beate Scherf (1993) show that there are many poorly described breeds of chicken on the verge of extinction that are kept by traditional societies. A broader view of the origins and distribution of chicken breeds in the world should help us to understand more clearly their role in both subsistence and the interaction of human cultures.

Roger Blench and Kevin C. MacDonald

Note

1. The authors are grateful to Zev Handel and James Matisoff for providing unpublished documentation of chicken names in Sino-Tibetan languages.

Bibliography

- Astill, G., and A. Grant. 1988. *The countryside of medieval England*. Oxford.
- Bellwood, Peter. 1995. Austronesian prehistory in Southeast Asia: Homeland, expansion and transformation. In *The Austronesians: Historical and comparative perspectives*, ed. Peter Bellwood, James J. Fox, and Darrell Tryon, 96–111. Canberra, Australia.
- Benecke, N. 1993. On the utilization of domestic fowl in central Europe from the Iron Age up to the Middle Ages. *Archaeofauna* 2: 21–31.
- Blench, R. M. 1995. A history of domestic animals in north-eastern Nigeria. *Cahiers des Sciences Humaines* 31: 181–238. Paris.
- Brewer, D. J., D. B. Redford, and S. Redford. 1994. *Domestic plants and animals: The Egyptian origins*. Warminster, England.
- Buitenhuis, H. 1983. The animal remains from Tell Sweyhat, Syria. *Palaeohistoria* 25: 131–44.
- Carpenter, Lawrence K. 1985. How did the “chicken” cross the Andes? *International Journal of American Linguistics* 51: 361–4.
- Carter, G. F. 1971. Pre-Columbian chickens in America. In *Man across the sea: Problems of pre-Columbian contacts*, ed. C. L. Riley, J. C. Kelley, C. W. Pennington, and R. L. Rands, 178–218. Austin, Tex.
- Carter, H. 1923. An ostrakon depicting a red jungle-fowl. *Journal of Egyptian Archaeology* 9: 1–4.
- Coltherd, J. B. 1966. The domestic fowl in ancient Egypt. *Ibis* 108: 217–23.
- Crawford, R. D. 1984. Domestic fowl. In *Evolution of domesticated animals*, ed. I. L. Mason, 298–311. London.
- Darby, William J., Paul Ghalioungui, and Louis Grivetti. 1977. *The gift of Osiris*. 2 vols. London.
- Dmitriev, N. G., and L. K. Ernst. 1989. *Animal genetic*

- resources of the USSR. FAO Animal Production and Health Paper No. 65. Rome.
- Fumihito, A., T. Miyake, S.-I. Sumi, et al. 1994. One subspecies of the red junglefowl (*Gallus gallus gallus*) suffices as the matriarchic ancestor of all domestic breeds. *Proceedings of the National Academy of Sciences* 91: 12505-9.
- Hernandez-Carrasquilla, F. 1992. Some comments on the introduction of domestic fowl in Iberia. *Archaeofauna* 1: 45-53.
- Johnston, H. H. 1886. *The Kili-Manjaro expedition; a record of scientific exploration in eastern equatorial Africa*. London.
- Kuit, H. G., A. Traore, and R. T. Wilson. 1986. Livestock production in central Mali: Ownership, management and productivity of poultry in the traditional sector. *Tropical Animal Health and Production* 18: 222-31.
- Langdon, R. 1990. When the blue-egg chickens come home to roost. *Journal of Pacific History* 25: 164-92.
- Loftus, Ronan, and Beate Scherf. 1993. *Worldwatch list*. Rome.
- MacDonald, K. C. 1992. The domestic chicken (*Gallus gallus*) in sub-Saharan Africa: A background to its introduction and its osteological differentiation from indigenous fowls (*Numidinae* and *Francolinus* sp.). *Journal of Archaeological Science* 19: 303-18.
1995. Why chickens?: The centrality of the domestic fowl in West African ritual and magic. In *Animal symbolism and archaeology*, ed. K. Ryan and P. J. Crabtree, 50-6. Philadelphia, Pa.
- MacDonald, K. C., and D. N. Edwards. 1993. Chickens in Africa: The importance of Qasr Ibrim. *Antiquity* 67: 584-90.
- Manessy, G. 1972. Les noms d'animaux domestiques dans les langues voltaïques. *Langues et techniques, nature et société; approches linguistiques*, Vol. 1, ed. J. M. C. Thomas and L. Bernot, 301-20. Paris.
- Moiseeva, I. G., Zhang Yuguo, A. A. Nikiforov, and I. A. Zakharov. 1996. Comparative analysis of morphological traits in the Mediterranean and Chinese chicken breeds: The problem of the origin of the domestic chicken. *Russian Journal of Genetics* 32: 1349-57.
- Mudida, N., and M. Horton. 1996. Subsistence at Shanga: The faunal record. In *Shanga: The archaeology of a Muslim trading community on the coast of East Africa*, ed. M. Horton, 378-93. Nairobi.
- Nishi, Y. 1990. Can fowls fly hundreds of miles over the Himalayas? In *Asian languages and general linguistics*, ed. T. Sakiyama and A. Sato, 55-77. Tokyo.
- Plug, Ina. 1996. Domestic animals during the early Iron Age in southern Africa. In *Aspects of African archaeology: Papers from the 10th Congress of the Pan-African Association for Prehistory and Related Studies*, ed. G. Pwiti and R. Soper, 515-20. Harare, Zimbabwe.
- Shmidt, P. 1927. *Domashnyaya ptitsa i ee istoriya* (Poultry and its history). Moscow.
- Skinner, N. A. 1977. Domestic animals in Chad. In *Papers in Chadic linguistics*, ed. P. Newman and R. M. Newman, 175-98. Leiden, The Netherlands.
- Taran, M. 1975. Early records of the domestic fowl in Judea. *Ibis* 117: 109-10.
- Thesing, R. 1977. *Die Großentwicklung des Hausbubns in vor- und frühgeschichtlicher Zeit*. Ph.D. dissertation, Munich University.
- West, B., and B.-X. Zhou. 1988. Did chickens go north? New evidence for domestication. *Journal of Archaeological Science* 15: 515-33.
- Williamson, K. In press. Did chickens go west? In *The origins and development of African livestock*, ed. R. M. Blench and K. C. MacDonald. London.
- Wood-Gush, D. G. M. 1959. A history of the domestic chicken from antiquity to the 19th century. *Poultry Science* 38: 321-6.
- Zeuner, F. E. 1963. *A history of domesticated animals*. London.
- Zorc, R. D. P. 1994. Austronesian culture history through reconstructed vocabulary (overview). In *Austronesian terminologies: Continuity and change*, ed. A. K. Pawley and M. D. Ross, 541-94. Canberra, Australia.

II.G.7 🐔 Chicken Eggs

*Hickety Pickety, my black hen
She lays eggs for gentlemen
Gentlemen come here every day
To see what my hen doth lay*

Anonymous

History

Eggs from many species of fowl have doubtless been consumed since the very beginning of humankind's stay on earth. In historical times, ancient Romans ate peafowl eggs, and the Chinese were fond of pigeon eggs. Ostrich eggs have been eaten since the days of the Phoenicians, whereas quail eggs, as hard-cooked, shelf-stable, packaged products, are now featured on many gourmet food counters in the United States and Japan. Other eggs consumed by various ethnic groups include those from plovers, partridges, gulls, turkeys, pelicans, ducks, and geese. Turtle eggs have been highly prized, and in starvation situations, any eggs, even those of alligators, have been relied upon.

In this chapter, however, only avian eggs (and these mostly from the chicken) are discussed. Avian eggs in themselves constitute a huge subject: In 1949, A. L. Romanoff and A. J. Romanoff published a book in which they attempted to compile all the facts known, at the time, about the eggs of birds. It contained over 2,400 reference citations.

It is almost obligatory in writing about eggs to first deal with that age-old question: Which came first, the chicken or the egg? Those who believe in creationism rely on holy books, like the Bible, which specify that animals were created. Thus, the chicken came first. But, as Harold McGee has pointed out, the eggs of reptiles preceded by far the evolution of the first birds; consequently, "[e]ggs . . . are millions of years older than birds." He added that "*Gallus domesticus*, the chicken, more or less as we know it, is only 4 to 5 thousand years old, a latecomer even among the domesticated animals" (1984: 55).

McGee placed the ancestors of *Gallus domesticus* (as the Romans named it) in Southeast Asia or India.

Maguelonne Toussaint-Samat was a bit more specific in writing that the chicken is "a descendant of a bird from the Malaysian jungle" (1992: 351). Still others have designated Burma and quite recently Thailand as its homeland, and there has been an argument for multiple origins with as many as four different species of jungle fowl contributing to the modern chicken (Smith and Daniel 1982; Fumihito et al. 1994; Simoons 1994).

The chicken was not only a tardy arrival as a food animal; it may have been domesticated not for food purposes at all but rather because of a perceived need for an on-hand supply of birds for sacrifice and divination. Frederick Simoons (1994) pointed out that these have been traditional uses of the chicken in Southeast Asia and are still roles the bird is called upon to fill by many peoples of the region.

Another early, but nonfood, use for the chicken in Southeast Asia was the sport of cockfighting – a sport that to this day remains immensely popular there, despite thousands of years of opposition by religions such as Hinduism and Buddhism (Simoons 1994).

But cockfighting was not confined to Southeast Asia, and as interest in it spread, so did the chicken. Although there was mention of the bird in Egypt in the early fourteenth century B.C., there seems to be no subsequent mention of it for many centuries thereafter, prompting speculation that it may have disappeared after its initial introduction (Smith and Daniel 1982). The chicken, however, was also in Persia at an early date, and from there it (along with cockfighting) spread out into ancient Greece, where it joined ducks, geese, and guinea fowl in poultry yards around the fifth century B.C. (Toussaint-Samat 1992).

Although a few recipes from the Greeks indicate that eggs were used for baking, and the physicians who compiled the Hippocratic Corpus recommended lightly cooked eggs as the most nourishing way to prepare them, and Aristotle systematically opened eggs at different points in their incubation to describe what he saw, it is doubtful that egg production was, initially at least, a very important reason for maintaining the bird (Smith and Daniel 1982; Toussaint-Samat 1992).

In the Roman period, however, although cockfighting remained a primary reason for keeping chickens, eggs were finally beginning to catch on in the kitchen. The recipes of Apicius (25 B.C.) reveal custards, omelets, and eggs used in a variety of other dishes, as well as by themselves in hard-boiled form (McGee 1984; Toussaint-Samat 1994). The Roman physician Galen, however, condemned fried eggs, saying they "have bad juice and corrupt even the foods mixed with them" (cited in Smith and Daniel 1982: 366).

Yet in other parts of the world at about this time, it would appear that egg consumption was avoided more often than not. In part this was because eggs (as well as chickens) were used for divination, in part because eggs were regarded as a filthy food (the product of the semen of a cock), in part because of food taboos (such as those that regulated the diets of preg-

nant women and their youngsters), and often because it was believed wasteful to eat the egg instead of waiting for the chicken.

China, however, constituted a very large exception to egg avoidance. Both chickens and eggs were important sources of animal protein, and the Chinese are said to have encouraged their use throughout the rest of East Asia. In Southeast Asia and in the Pacific Islands (where the chicken was distributed early on by Asian colonizers), as well as in China, a taste was developed for brooded eggs (with well-developed fetuses). In addition, the Chinese became partial to "100-year-old" eggs that tend to make Westerners gag. But despite the name, the eggs are buried for only a few months in a mixture of saltpeter, clay, tea leaves, and other materials that cause the shells to turn black. The interiors of such eggs take on a "hard-boiled" appearance with green veins running through them (Toussaint-Samat 1992; Simoons 1994).

Moving from the East back to the West after the fall of Rome, darkness descended on both Europe and the chicken, and little is known about the use of eggs on the Continent until the sixteenth century. In the meantime, the Iberians had discovered the New World, triggering many debates for scholars – among them the question of whether chickens were on hand in the Americas when Columbus first arrived in 1492. The answer seems to be no, at least as far as the West Indies (where the Spaniards introduced them) are concerned, but may very possibly be yes, as far as South America is concerned. There, the native people had names of their own for a kind of chicken – and the only one in the world that lays blue and green eggs (Smith and Daniel 1982). Page Smith and Charles Daniel have speculated that this South American *araucana* was not an "indigenous chicken" but instead the product of a union of American grouse with Asian chickens that reached South America with earlier, unrecorded voyagers (1982: 31), most likely Pacific Islanders.

European literature is not completely silent on eggs prior to the Renaissance; we are warned in *Don Quixote de la Mancha* (by Miguel de Cervantes [1547-1616]) not to put them all in one basket, and Francis Bacon (1561-1626) commented on those who would burn their houses down just to roast an egg. Throughout the preceding Middle Ages, however, eggs were mostly mentioned in nursery rhymes. In large part, one suspects that this considerable silence was because eggs were classified as meat by the Church; what with Fridays, Lent, and numerous other days when meat was proscribed, eggs were off the menu about half of the days of the year. As a consequence, most eggs were either used for hatching or were saved to be eaten at Easter. To keep them from spoiling, they were dipped in either liquid fat or wax and then decorated to make them more attractive – hence the custom of "Easter eggs" (Toussaint-Samat 1992).

Although some French and English recipe books from the late fourteenth century contain directions for making custards, omelets, and baked eggs (McGee 1984), Smith and Daniel (1982) suggest that the renaissance of the chicken came only in the sixteenth century with the work of Ulisse Aldrovandi (1522–1605), an Italian who wrote nine volumes on animals, including one on chickens. It was the dawn of a new age of science, and as Aristotle had done some 1,800 years earlier, Aldrovandi systematically examined the egg while at the same time adroitly dodging “that trite and thus otiose . . . question, whether the hen exists before the egg or vice versa” (quoted in Smith and Daniel 1982: 45). It might have been a new age of science, but the Church still had a firm answer for this old riddle.

The egg had made considerable culinary headway by 1651, when Pierre François de la Varenne published *Le cuisinier françois*, a cookbook that provided 60 recipes for eggs (Tannahill 1989). But it is the era embracing the eighteenth and early nineteenth centuries that has been characterized as “The Century of the Chicken” (Smith and Daniel 1982) because of the considerable amount of scientific interest the bird generated. Upon learning about elaborate hatching ovens in Egypt, the French naturalist René de Reaumur wrote a treatise on the subject, and breeding chickens became a preoccupation of European (and North American) country squires with a scientific turn of mind (McGee 1984).

This effort was given considerable impetus in the nineteenth century with the opening of the Chinese port of Canton (in 1834) to foreign traders. One of the first English vessels in the new trade returned with a few chickens of a Chinese breed – “Cochin” fowl, as they were ultimately called – as a present for Queen Victoria. In addition to their startlingly spectacular appearance, the Cochin chickens were superior in meat and egg production to established Mediterranean and European breeds. When they were first exhibited in England, tens of thousands of people showed up to stare, and all breeders had to have one. The same phenomenon took place at the Boston Poultry Show of 1849, where the birds again attracted crowds in the thousands, and even Daniel Webster attended to heap praise on the fowl.

Chickens had suddenly gained a new prominence among barnyard animals, and others of the great Asian breeds followed the Cochin to the West to perpetuate the chicken craze (Smith and Daniel 1982; McGee 1984). Breeding for show – with major emphasis on feather coloring and comb type – continued throughout much of the rest of the nineteenth century, with over 100 different breeds and color variations the result.

The “Century of the Chicken” was also a century of the egg, during which it was incorporated into diets as never before. Boiled eggs for breakfast became a favorite of many, and it was said that entire families of

Parisians crowded around every Sunday to admire the dexterity of their sovereign, Louis XIV, who could knock off the small end of an egg with his fork in a single stroke (Toussaint-Samat 1992). The king was also very fond of meringues. Cookbooks provided careful instructions for the preparation of omelets and the poaching of eggs; mayonnaise was invented in the middle of the eighteenth century; the Americans followed the English example of marrying bacon with eggs; and the baking industry boomed (Trager 1995). Consequently, as the end of the nineteenth century approached, eggs were very much in demand in the West, and the emphasis in chicken breeding shifted from show-bird characteristics to productive capacity for eggs, or meat, or both.

Early in the twentieth century, Artemus Ward, in *The Grocer's Encyclopedia*, indicated something of the way that demand was being met. He wrote of “large poultry farms [where] eggs are produced and handled very much as the product of any other factory . . . but,” he added, “the greater part of the country's egg supply is still represented by accumulations from thousands of general farmers scattered all over the country” (1911: 223).

This is hardly the case today. Poultry sheds have become meat and egg factories with automated hatcheries (McGee 1984). Major steps in this direction took place in the 1930s and 1940s as John Tyson pioneered the vertical integration of the poultry industry. In 1956, the animal health-products division of Merck Pharmaceutical Company began production of a drug that prevented flock-destroying epidemics of coccidiosis. These events were accompanied by the development of high-protein feeds, after which chicken and egg production became truly automated industries (Trager 1995).

The egg is a fine (and cheap) source of high-quality protein, as well as iron, vitamins E and B₁₂, folacin, riboflavin, and phosphorous, and was long regarded as a near-perfect food. But the egg's yolk is also a source of considerable cholesterol, and with the implication of this substance in blocking heart arteries, demand for fresh eggs fell by almost 25 percent in the decade of the 1980s. In addition, eggs have been blamed of late (with some regularity) for outbreaks of salmonellosis. But although the use of fresh eggs has fallen off, the sale of food products containing eggs has risen significantly (Mergen et al. 1992). The industry is hardly a dying one.

The Shell Egg Industry

The egg industry in most of the world is based on chicken eggs. In Southeast Asia there is also a duck egg market. As ethnic populations move, the food products they desire move with them. With the rather large number of Southeast Asian natives now living in other parts of the world, a geographically widespread demand for duck eggs has become a relatively recent phenomenon.

Table II.G.7.1 *Egg production of several countries of the world (in millions of eggs)*

Country	1969	1979	1989
United States	68,700	69,228	67,042
Former Soviet Union	37,000	65,585	84,600
Japan	27,565	33,150	40,383
Germany ^a	15,036	14,962	12,275
United Kingdom	18,450	18,555	17,834
France	11,200	13,840	15,150
Italy	10,281	11,037	11,223
Brazil	7,140	11,035	10,140
Spain	9,670	7,200	12,174
Poland	6,400	8,670	8,200
Mexico	5,657	10,390	14,700
Canada	5,655	5,551	5,500
Netherlands	4,370	8,187	10,711
Belgium, Luxembourg	3,900	3,453	2,724
Czechoslovakia	3,410	4,225	5,628
Romania	3,200	7,085	7,600
Argentina	2,940	3,350	3,350
Hungary	2,600	4,450	4,250
Australia	2,380	3,395	3,286
Republic of South Africa	2,028	3,083	4,012
Turkey	-	-	7,200
China	-	-	140,900
Korea	-	-	6,919

Source: USDA (1970, 1980, 1990).

^aProduction for East and West Germany was combined.

Table II.G.7.1, which includes only the markets for chicken eggs, lists the countries leading in egg production. Data on egg production in China, prior to 1989, are not available, but production there has always been extensive. Prior to 1940, China was the leading nation in the export of dried egg products. In Turkey and Korea, egg production has recently become significant. Figures for Eastern European countries for the 20-year period 1969 to 1989 show an increase in egg production in most of the countries, especially in the former Soviet Union. In countries with production controls, such as Canada, Australia, and the United Kingdom, there was almost no increase. The lower production in the United Kingdom for 1989 was likely influenced significantly by bad publicity relative to the safety of eggs during the mid-1980s. Overall, most countries listed in Table II.G.7.1 had an increase in egg production in excess of population increase.

The concentration of egg production has moved rapidly to larger units in the United States. In 1959, over 49 percent of all eggs in the United States were produced by flocks of less than 1,600 hens. By 1974, only 5.44 percent of all eggs were produced by such small flocks. In 1990, less than 1,000 companies produced over 97 percent of all eggs, with the smallest of these commercial farms having over 30,000 layers.

Egg consumption figures for countries around the world can be approximated by dividing human popu-

lation figures for each country into the production figures listed in Table II.G.7.1. There is an international trade for eggs, but in most countries it is a relatively small percentage of total production. The Netherlands might be an exception, as they export many eggs to other European Economic Community (EEC) countries. Shell egg consumption in the United States has been declining since about 1950, when annual consumption was about 400 eggs per person with over 95 percent being in shell egg form. The 1990 consumption of shell eggs had decreased to about 200 per capita, with egg products accounting for over 40 more eggs per person. The use of egg products and eggs in prepared form has been increasing by 1 to 2 percent annually for the last several years. Publicity about potential danger from bacteria in shell eggs will likely result in a continuation of the shift to pasteurized egg products, particularly by institutional feeding establishments.

The egg products industry in the United States is expanding in numbers of companies involved and in volume from each company. The basic products are liquid, frozen, or dried whole egg, egg white, and egg yolk. For reasons of convenience, the shift has been from frozen and dried to liquid products. Developments in pasteurizing procedures and in the aseptic handling of liquids suggest that the shift to liquid will accelerate.

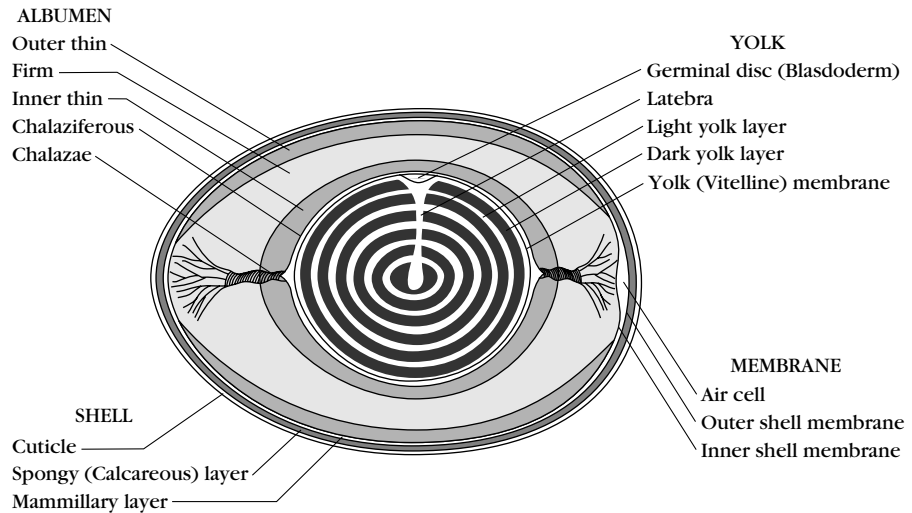
The egg industry is in a growth phase worldwide and will continue to expand because of the wide acceptance of eggs as food of a high nutritional value, especially in terms of protein quality, which makes the egg such a valuable food source for the expanding world population. Research and technology are striving to overcome the problems presented by cholesterol content and microbiological safety.

Egg Formation

At the time of hatching, the female chicken has a rudimentary ovary with over 2,000 ova or immature egg yolks called oocytes. The egg yolk develops while in a follicle of the ovary. The remainder of the egg, especially the white, or albumen, and the shell, is formed in the oviduct. From these values it is apparent that major changes in the size of the reproductive system take place in relatively short time periods.

It is generally acknowledged that a hen forms an egg in about two weeks. This is true except for the very small core of the yolk. The yolk of the egg is formed in three stages: (1) the part formed during embryonic development of the female chick; (2) the normal slow development of the ovum from the time of hatching to a point in sexual maturity some 10 days prior to ovulation; and (3) the accelerated growth period during the last 10 days before ovulation (release of the ovum or yolk into the oviduct) (Stadelman and Cotterill 1995). The yolk increases in size during the rapid growth stage by the deposition

Figure II.G.7.1. Structure of the chicken egg.



of layers of yolk material. The concentric rings of growth remain intact until the yolk membrane, the vitelline egg membrane, is broken. The yolk of a normal chicken egg makes up 27 to 33 percent of the total egg weight.

In the oviduct the albumen, or white portion, of the egg, which accounts for about 60 percent of the total egg weight, is deposited around the yolk in about 4.5 hours. The oviduct is made up of five distinct regions. The yolk is released from the ovarian follicle into the infundibulum. A small amount of thick albumen rich in mucin fibers is deposited around the yolk during its 20-minute time of passage, due to peristaltic action, into the magnum section of the oviduct. In the magnum section, the yolk collects the rest of the thick albumen and some thin albumen. The developing egg spends about 3 hours in the magnum section prior to entering the isthmus section where the shell membranes and remaining thin white are added during about a 1-hour stay.

The edible portion of the egg is now complete, but about 19 hours is spent in the uterus section where the shell is deposited over the shell membranes. In the posterior section of the uterus the pigments of colored egg shells are deposited on the shell just prior to movement of the intact egg through the vagina.

From the time the first thick albumen is deposited on the yolk until the shell membranes are in place, the yolk spins due to the peristaltic movements of the oviduct. The spinning motion causes mucin fibers of the thick albumen to be twisted on opposite sides of the yolk into the chalaza. The chalaza forms a tight mesh over the yolk surface with the loose ends enmeshed in the layers of thick albumen. The chalaza aids in keeping the yolk centered in the albumen.

Another part of the egg often observed is the air cell. This forms between the inner and outer shell

membranes after the egg is laid and is usually found in the large end. At the time of lay, the egg is at the body temperature of the hen, about 107° F (42° C), but as it cools, the volume shrinks, which starts the air cell. During the time the egg is held, there is a continual loss of moisture from the egg albumen, which results in an ever increasing air cell size. The rate at which moisture is lost is controlled by atmospheric humidity and porosity of the shell.

Egg Structure

The parts of the egg are shown in Figure II.G.7.1. As discussed in the section on the formation of the egg, the ovum starts development during the embryonic growth of the female chick. The female germ cell is in this structure. During the rapid growth phase of egg formation, the latebra grows to keep the germinal disk on the surface of the yolk. The vitelline membrane expands quickly during the rapid growth phase of yolk development. In a fully developed yolk the vitelline membrane is about 0.024 millimeter (mm) thick (Needham 1931). Depending on the methods used in histological examination of the membrane, there are either two or three layers (Romanoff and Romanoff 1949).

The yolk is deposited in concentric layers during formation. In hard-cooked egg yolks it is possible to differentiate light and dark rings when highly pigmented feedstuffs are fed for a limited period of time each day. The size of the yolk varies directly with egg size. The percentage of the total egg in the yolk varies from 27 to 33 percent. The percentage of the egg as yolk varies with egg size (Marion et al. 1964), with strains laying smaller eggs having a higher percentage of the egg as yolk. According to F. E. Cunningham, O. J. Cotterill, and E. M. Funk (1960), as hens age, the percentage of yolk in the egg increases. Cotterill and

G. S. Geiger (1977) studied yield records from the Missouri Random Sample test and found that the yolk size in eggs decreased from 1965 to 1975. W. W. Marion and others (1964) suggested that all deviations in egg component part percentages are covariates of egg size.

The size of the yolk increases slowly during storage due to a very slow balancing of osmotic pressures between the yolk and albumen. A series of papers by J. Needham and M. Smith (1931), Smith and J. Shepherd (1931), Smith (1931), J. Needham, M. Stephenson, and D. M. Needham (1931), and J. Needham (1931) reported on the relations of yolk to albumen in the hen's egg. The vitelline membrane is a unique entity in its ability to maintain about 50 percent solids in the yolk, with only 13 percent solids in the albumen over many months of storage. At this time, a proven explanation for this phenomenon is not available.

The albumen of the hen's egg consists of four layers that can be identified in the broken-out freshly laid egg. The location of the layers in the intact egg is indicated in Figure II.G.7.1. The inner thick, or chalaziferous, layer is next to the vitelline membrane. This layer of albumen is rich in mucin fibers that are generally lacking in the thin white layers. The outer thick layer is a complex of lysozyme-mucin proteins. This complex disintegrates over time when eggs are in storage.

The shell membranes, inner and outer, are composed of keratin fibers arranged in a web pattern to form a barrier to microbial invasion of the egg contents. The inner membrane is about one-third the thickness of the outer membrane (Romanoff and Romanoff 1949). The combined thickness of the two membranes is about 0.1 mm. Passage of gases or liquids through the membranes occurs largely by osmosis or diffusion.

The egg shell must meet a number of requirements. It must be strong enough and rigid enough to withstand the weight of the adult hen. It must be porous enough to allow respiration for the developing embryo during incubation and still compact enough to prevent microbial invasion or the escape of too much moisture. Additionally, the shell serves as a supply of minerals for the nutrition of the embryo. The outer surface of the egg is the shell, which consists of four layers, all composed primarily of calcium carbonate. The thickness of the egg shell varies among hens. It is normally thickest for hens just starting egg production; thickness decreases as the hens extend their egg-laying cycle. A hot environmental temperature also results in thinner egg shells, as does a low level of calcium in the hen's diet.

Egg Quality Evaluation and Preservation

The basis for quality determination of eggs by nondestructive means in the United States is the *Egg Grading Manual* (U.S. Department of Agriculture 1977). The same characteristics laid out in the manual are used in all countries with various degrees of emphasis

on the several quality factors. Quality determinations are divided into external and internal factors. External quality factors are soundness of the shell, cleanliness, and egg shape. The internal factors are air cell size, albumen viscosity, and yolk shadow. The internal factors are judged by passing the egg in front of a bright light source. Other internal factors are freedom from blood spots, bloody albumen, meat spots, or other inclusions. Equipment is currently available that allows for quality evaluation by candling at rates in excess of 72,000 eggs per hour.

In 1981, the quality standards for grades of eggs in the United States were modified (U.S. Department of Agriculture 1981). The grades of eggs that can be offered for sale at retail are AA, A, B, and B*, referred to as B star. The requirements for each grade are detailed for each external and internal quality factor. The application of these standards are discussed by W. J. Stadelman and Cotterill (1995).

Numerous laboratory methods for quality evaluation of eggs have been prepared. Most of these methods are destructive in that the shell is broken and measurements are made on the liquid contents. The most widely accepted method is the Haugh Unit for expressing the albumen condition. This measurement was presented by R. R. Haugh (1937) and modified by A. W. Brant, A. W. Otte, and K. H. Norris (1951). A lesser used system is the United States Department of Agriculture (USDA) score as suggested by Brant and colleagues (1951), which attempts to correlate visual appearance of the broken-out egg, Haugh Units, and candled grade.

Evaluation of shell quality is on the basis of shell strength. The breaking strength has been found to be closely related to shell thickness. On broken-out eggs, measurement of shell thickness is a common method. As the specific gravity of freshly laid eggs is determined primarily by shell thickness, a nondestructive estimation of shell thickness can be made by determining specific gravity of the intact egg.

In terms of quality preservation, the most frequently considered item is albumen condition, which is often expressed as Haugh Units. For a high-quality or grade AA egg, the Haugh Units should be above 78. This value lowers over time as the thick albumen thins. The rate of albumen thinning is a function of temperature. The breakdown or thinning of albumen is relatively rapid at high temperatures, 40° Celsius (C), and slows to almost no change at 1° C. Other than temperature, the carbon dioxide content of the atmosphere surrounding the egg affects the rate of carbon dioxide loss from the egg. With the loss of carbon dioxide, the pH of the albumen rises from about 7.6 in a fresh egg to 9.7 in a stale egg.

Humidity of the atmosphere influences rate of water loss from the egg, which results in increased air cell size. Ideal conditions for long-term storage of eggs are a temperature between 1° C and 3° C and a relative humidity of about 80 percent. For long-term stor-

age, egg shells are usually coated with a colorless, odorless mineral oil that seals the pores of the shell.

A frequently neglected consideration in quality preservation is maintaining cleanliness and soundness of the shells. In summary, egg quality preservation is a function of time, temperature, humidity, and handling.

Egg Sizes

In most parts of the world egg sizes are based on metric weights, with each size up or down being with a 5-gram grouping. Eggs are sold by the dozen in the United States. Egg sizes are based on weight per dozen ranging from pee wee (15 ounces [oz]), small (18 oz), medium (21 oz), large (24 oz), extra large (27 oz), to jumbo (30 oz). For consumer satisfaction it is desirable to pack cartons holding a dozen with uniform-sized eggs of the appropriate weight. Many of the jumbo-sized eggs will contain two yolks.

Egg Chemistry

Shell eggs consist of about 9.5 percent shell, 63 percent albumen, and 27.5 percent yolk, according to Cotterill and Geiger (1977). The total solids of the albumen, yolk, and whole egg are about 12 percent, 52 percent, and 24 percent, respectively. Reviews of the chemistry of egg components were written by R. E. Feeney (1964), T. L. Parkinson (1966), Stadelman (1976), and W. D. Powrie and S. Nakai (1990). Extensive analytical data on the nutrient composition of eggs are given in a publication by the U.S. Department of Agriculture (1989).

The shell membranes are composed of protein fibers, mostly keratin with some mucin, arranged in net fashion to form a semipermeable membrane. The shell membranes provide a significant barrier to bacterial invasion of the albumen.

The egg white consists of a number of proteins, a small amount (less than 1 percent) of carbohydrates and minerals, and no lipids. The composition of the albumen of a freshly laid egg is about 87 percent moisture, 11.5 percent protein, 0.8 percent ash, and 0.7 percent carbohydrates. The carbohydrates are the sugars in glycoproteins. The predominant protein is ovalbumin. The complete amino acid sequence with 385 residues has been determined by A. D. Nisbet and others (1981). The molecular weight of the polypeptide chain is 42,699.

The pH of the albumen in a freshly laid egg is about 7.6, which increases during storage to as high as pH 9.7. The rate of pH change is influenced by temperature, air movement, and shell quality. The pH increase is the result of carbon dioxide loss from the albumen. The pH of the albumen depends on the equilibrium between the dissolved carbon dioxide, bicarbonate ion, carbonate ion, and proteins.

The protein ovotransferrin is referred to as conal-

bumin in older literature. Its unique characteristic is its ability to bind multivalent metallic ions. With aluminum salts a white precipitate is formed; with copper, the flocculant is yellow; and with ferric iron a red color results. The ovotransferrin is one of the most heat-labile proteins of the albumen. Powrie and Nakai (1990) discuss methods for the isolation of each of the proteins.

Egg yolk might be described as a complex system containing a variety of particles suspended in a protein solution. Another description is that it is an oil-in-water emulsion, with lecithin and the lysoproteins aiding in maintaining a stable emulsion. The yolk consists of a plasma with suspended granules. Fresh egg yolk contains about 52 percent solids and has a pH of 6.0, which rises during storage to 6.9.

Nutritional Value

Eggs are a popular food in all countries of the world and have been since ancient times. Before agriculture developed, eggs were gathered from birds' nests for human food. Although eggs contain about 75 percent water, they are a rich source of high-quality protein and are often used as the protein against which other protein sources are compared. Eggs are also important sources of unsaturated fatty acids, iron, phosphorus, trace minerals, vitamins A, E, and K, and all B vitamins. As a natural source of vitamin D, eggs rank second only to fish oils. Eggs are low in calcium, as the shell is not eaten, and they are devoid of vitamin C (Stadelman and Pratt 1989). The high nutrient density of eggs relative to their caloric content makes them an excellent food for many people with special dietary needs (Stadelman et al. 1988).

Much has been written concerning the relationship of plasma cholesterol level to coronary problems in humans, yet some individuals do not adequately differentiate between plasma, or serum, cholesterol and dietary cholesterol. For most people it would seem that there is only a very slight relationship between the level of cholesterol in their diet and the serum cholesterol level in their blood.

Microbiology

The egg contents at the time of laying are generally free of microbial contamination. During the 1980s it was found that a few eggs, estimated at 1 in 20,000, might contain a bacteria, *Salmonella enteritidis*, in the yolk at the time of production. This bacteria is of great concern to health officials and the egg-producing industry because the organism can cause food poisoning in humans.

The egg has a number of barriers to bacterial invasion. The shell and shell membranes act as physical barriers. In the albumen there are several proteins that influence the ability of organisms to colonize in the

egg. Lysozyme will digest cell walls of some bacteria. Avidin removes biotin from the available nutrients for bacteria, and ovotransferrin chelates with ions of iron, making it unavailable. When eggs are cooked, the tying up of biotin or iron is eliminated so these materials are readily available for microorganisms or humans.

Packaging

As egg shells are subject to cracking, packaging has been developed to minimize this loss. In early days eggs were loosely packed in rigid containers using whole grain oats to keep the eggs from contacting the sides of the container or each other. A step forward in innovation was the use of fillers and flats made of cardboard to keep eggs in individual cells in wooden cases. The wooden case was standardized to hold 30 dozen eggs. The next move was to fiberboard cases and a pulp paper filler flat. These cases are now in use for domestic shipments. Most export shipments are still in wooden cases using the filler flat.

For the retailing of eggs, many packages have been and are still being used. The poorest of these, as far as protecting the egg, is a bulk display with paper bags to carry a dozen or more eggs. Pulp paper cartons have been developed to hold from 2 to 18 eggs. These may be coming back into use because of environmental concerns regarding plastic foam cartons.

The Liquid Egg Industry

Shell eggs are converted to liquid products by the removal of the shell. It is required in the United States and some other countries that the shells be removed with no commingling of the shell and liquid portions. The liquid content of the egg is handled as albumen, yolk, whole egg, and various blends of yolk and albumen. Commercial equipment can break and separate yolks from albumen at rates in excess of 36,000 eggs per hour.

Processing Liquid Eggs

The steps of processing liquid eggs include pasteurization, homogenization, packaging, and refrigeration. The liquid products may be converted to frozen, dehydrated, or dried products. An excellent historical review covering early development of the United States egg-products industry was prepared by J. W. Koudele and E. C. Heinsohn (1960).

The pasteurization of eggs is accomplished by heating the liquid to a sufficiently high temperature for a long enough time to kill any pathogenic microorganisms that might be present. As egg proteins coagulate when heated, the pasteurization temperatures and times must be carefully controlled in order to obtain the bacterial reduction with minimal damage to the functional properties of the egg product. Heat damage can be minimized by maintaining a high degree of turbulence in the egg product during the pasteurization process. For whole eggs in the

United States, a temperature of 60° C (140° F) for a minimum of 3.5 minutes is required. Higher temperatures for a shorter holding time are accepted. C. R. Murdock and others (1960) compiled information on minimal accepted pasteurization times and temperatures in several countries. In Poland whole egg pasteurization requires 66.1° C to 67.8° C for 3 minutes, in China 63.3° C for 2.5 minutes and in Australia 62.5° C for 2.5 minutes.

Pasteurization temperatures for albumen are sometimes lower than for the whole egg, using longer holding times. The addition of aluminum salts to the albumen results in chelation, with the ovotransferrin allowing the use of higher pasteurization temperatures for albumen (Lineweaver and Cunningham 1966). Ovotransferrin is the most heat-labile of all proteins in the albumen. Egg yolk is rather heat-stable and can be pasteurized by using slightly higher temperatures or longer times than for the whole egg. Cunningham (1990) reported detailed procedures for the pasteurization of different liquid egg products.

A procedure for the ultrapasteurizing of liquid whole eggs was described by H. R. Ball, Jr., and colleagues (1987), and by K. R. Swartzel, Ball, and M. Hamid-Samimi (1989). They utilized temperatures that are higher than the minimum required in the United States with high turbulence in the holding tubes and projected a usable shelf life of the ultrapasteurized whole egg products of up to 24 weeks when stored at 4° C or lower. The commercialization of this procedure will reduce the need for frozen products.

The homogenization of liquid whole egg may be done either before or after pasteurization. It usually follows pasteurization so that any clumping of materials due to the heating would be dispersed in the homogenization operation.

The packaging of egg products is an ever changing operation. The early processors, from the late nineteenth century until the mid-twentieth century, used metal cans, each containing about 30 pounds of product. The next package was a rigid plastic bucket of either 10- or 30-pound capacity. This was followed by the use of waxed cardboard boxes of 4- or 5-pound capacity. Small quantities were packaged in half-pint cardboard containers in a futile attempt at selling the egg products in retail food stores. At present the industry is moving toward the use of flexible film packages with an aseptic packaging technology.

Before the introduction of ultrapasteurization and aseptic packaging, the shelf life of liquid eggs was about 6 days. But now, with a predicted shelf life of up to 24 weeks (Swartzel et al. 1989), it may be possible to sell liquid product at retail. In earlier years much of the liquid product was packaged as indicated and frozen for distribution to food manufacturers.

During World War II there was a great demand for dried egg products for military feeding and to supply

civilian populations of countries in the war zones. The usual method of drying was by using a spray drier. Small amounts were pan-dried and freeze-dried. In the drying of eggs it was necessary to remove the reducing sugars normally present in egg albumen to minimize the discoloration that would result from a reaction between such sugars and amino acids abundant in the egg liquid. Desugaring is accomplished by a controlled bacterial or yeast fermentation, by a spontaneous microbial fermentation, or by an enzymatic fermentation using glucose oxidase.

W. M. Hill and M. Sebring (1990) reviewed methods for desugaring liquid eggs. D. H. Bergquist (1990) outlined procedures for the dehydration of liquid eggs and listed seven advantages of the dried products. These are as follows:

1. A low storage cost.
2. A low transportation cost.
3. The ease of sanitary handling.
4. The lack of susceptibility to bacterial growth during storage.
5. An allowance for more precise formulation.
6. The uniformity of product.
7. An allowance for development of new convenience foods.

During the last 20 years a number of additional processed egg products have been introduced in the marketplace, as described by Stadelman and others (1988). The most widely marketed to date have been hard-cooked egg products. They are available as peeled hard-cooked eggs, diced eggs, long eggs, and Scotch eggs. It was estimated that about 1 percent of all eggs sold in the United States in 1990 were in hard-cooked form. Generally, the hard-cooked egg sales are to restaurants rather than at retail.

In summary, eggs have long served as a food for humankind. They have a high-density nutritional value, especially with respect to the protein quality. They contain all nutrients required by humans except vitamin C, but as we do not utilize the shell, the egg is also deficient in calcium. Eggs are usually one of the first solid foods given to infants and frequently constitute a significant portion of the diet of the elderly. Valued for their versatility, they are employed to coagulate, to foam, to emulsify, to color, to flavor, and to control sugar crystallization in some candies. With this wide range of use, eggs remain one of our most economical food staples.

William J. Stadelman

Bibliography

- Ball, H. R., Jr., M. Hamid-Samimi, P. M. Foegeding, and K. R. Swartzel. 1987. Functional and microbial stability of ultrapasteurized, aseptically packaged refrigerated whole egg. *Food Science* 52: 1212-18.
- Bergquist, D. H. 1990. Egg dehydration. In *Egg science and technology*, ed. W. J. Stadelman and O. J. Cotterill, 285-323. Binghamton, N.Y.
- Brant, A. W., A. W. Otte, and K. H. Norris. 1951. Recommended standards for scoring and measuring opened egg quality. *Food Technology* 5: 356-61.
- Cotterill, O. J., and G. S. Geiger. 1977. Egg product yield trends from shell eggs. *Poultry Science* 56: 1027-31.
- Cunningham, F. E. 1990. Egg product pasteurization. In *Egg science and technology*, ed. W. J. Stadelman and O. J. Cotterill, 243-72. Binghamton, N.Y.
- Cunningham, F. E., O. J. Cotterill, and E. M. Funk. 1960. The effect of season and age of bird. 1. On egg size, quality and yield. *Poultry Science* 39: 289-99.
- Feeney, R. E. 1964. Egg proteins. In *Symposium on food: Proteins and their reactions*, ed. H. W. Schultz and A. F. Anglinier, 209-24. Westport, Conn.
- Fumihito, A., T. Miyake, S.-I. Sumi, et al. 1994. One subspecies of the red junglefowl (*Gallus gallus gallus*) suffices as the matriarchic ancestor of all domestic breeds. *Proceedings of the National Academy of Sciences* 91: 12505-9.
- Haugh, R. R. 1937. The Haugh Unit for measuring egg quality. *U.S. Egg and Poultry Magazine* 43: 552-5, 572-3.
- Hill, W. M., and M. Sebring. 1990. Desugaring egg products. In *Egg science and technology*, ed. W. J. Stadelman and O. J. Cotterill, 273-83. Binghamton, N.Y.
- Koudele, J. W., and E. C. Heinsohn. 1960. The egg products industry of the United States. *Kansas State University Agricultural Experimental Station Bulletin* No. 423.
- Lineweaver, H., and F. E. Cunningham. 1966. *Process for pasteurizing egg white*. U.S. Patent No. 3,251,697.
- Margen, Sheldon, et al., eds. 1992. *The wellness encyclopedia of food and nutrition*. New York.
- Marion, W. W., A. W. Nordskog, H. S. Tolman, and R. S. Forsythe. 1964. Egg composition as influenced by breeding, egg size, age and season. *Poultry Science* 43: 255-64.
- McGee, Harold. 1984. *On food and cooking*. New York.
- Murdock, C. R., E. L. Crossley, J. Robb, et al. 1960. The pasteurization of liquid whole egg. *Ministry of Labor and Health Services Monthly Bulletin* 19: 134-50.
- Needham, J. 1931. The relations between yolk and white in the hen's egg. V. The osmotic properties of the isolated vitelline membranes. *Journal of Experimental Biology* 8: 330-44.
- Needham, J., and M. Smith. 1931. The relations between yolk and white in the hen's egg. I. Introduction. *Journal of Experimental Biology* 8: 286-92.
- Needham, J., M. Stephenson, and D. M. Needham. 1931. The relations between yolk and white in the hen's egg. IV. The formation of lactic acid and alcohol by the yolk. *Journal of Experimental Biology* 8: 319-29.
- Nisbet, A. D., R. H. Saundry, A. J. G. Mair, et al. 1981. The complete amino acid sequence of hen ovalbumin. *Environmental Journal of Biochemistry* 115: 335-45.
- Parkinson, T. L. 1966. The chemical composition of eggs. *Journal of Science Food Agriculture* 17: 101-11.
- Powrie, W. D., and S. Nakai. 1990. The chemistry of eggs and egg products. In *Egg science and technology*, ed. W. J. Stadelman and O. J. Cotterill, 97-139. Binghamton, N.Y.
- Romanoff, A. L., and A. J. Romanoff. 1949. *The avian egg*. New York.
- Shenstone, F. S. 1968. The gross composition, chemistry and physicochemical basis of organization of the yolk and white. In *Egg quality, a study of the hen's egg*, ed. T. C. Carter, 26-64. Edinburgh.

- Simoons, Frederick J. 1994. *Eat not this flesh*. Second edition. Madison, Wis.
- Smith, M. 1931. The relations between yolk and white in the hen's egg. III. Gas exchange in infertile eggs. *Journal of Experimental Biology* 8: 312-18.
- Smith, M., and J. Shepherd. 1931. The relations between yolk and white in the hen's egg. II. Osmotic equilibration. *Journal of Experimental Biology* 8: 293-311.
- Smith, Page, and Charles Daniel. 1982. *The chicken book*. San Francisco.
- Stadelman, W. J. 1976. Egg proteins. In *Food colloids*, ed. H. D. Graham, 207-239. Westport, Conn.
- Stadelman, W. J., and O. J. Cotterill. 1995. *Egg science and technology*. Fourth edition. Binghamton, N.Y.
- Stadelman, W. J., V. M. Olson, G. A. Shemwell, and S. Pasch. 1988. *Egg and poultry-meat processing*. Chichester, England.
- Stadelman, W. J., and D. E. Pratt. 1989. Factors influencing composition of the hen's egg. *World's Poultry Science Journal* 45: 247-66.
- Swartzel, K. R., H. R. Ball, Jr., and M. Hamid-Samimi. 1989. Method for the ultrapasteurization of liquid whole egg products. U.S. Patent No. 4,808,425.
- Tannahill, Reay. 1989. *Food in history*. New York.
- Toussaint-Samat, Maguelonne. 1992. *A history of food*, trans. Anthea Bell. Cambridge, Mass.
- Trager, James. 1995. *The food chronology*. New York.
- U. S. Department of Agriculture. 1970, 1980, 1990. *Agricultural Statistics*. Washington, D.C.
1977. *Egg Grading Manual, Agricultural Handbook No. 75*. Washington, D.C.
1981. Revision of shell egg standards and grades. *Federal Register* 46 (149): 39566-73.
1989. *Composition of foods: Dairy and egg products, raw, processed, prepared*. Agricultural Handbook 8.1. Washington, D.C.
- Ward, Artemus. 1911. *The grocer's encyclopedia*. New York.

human owners. Captain James Cook noted in his journal of 1784 that dogs were fed and herded with the hogs. Titcomb (1969) relates that they were fattened on poi and that an informant remembered that relatives penned up puppies to keep them from eating unclean food. They were fed three times a day on poi and broth until fattened. Natives considered baked dog a great delicacy, but they never offered it to foreigners, who held dog meat in great abhorrence, according to the wife of a missionary in 1820.

Preparation of dogs as food is abstracted by Titcomb (1969), who writes: "Like the pig, the dog was killed by strangulation in order not to lose the blood. The blood was poured into a calabash (gourd) used for cooking, and red hot stones were added, causing the blood to coagulate; it was then ready to eat." The meat might be baked or cut up and wrapped in *ti* leaf bundles and cooked in boiling water. An informant stated that the upper part of the backbone was preferred by many diners, although some liked the ribs best, and others preferred the brain. The latter has been attested to by dog skulls, with holes in the crania, found in archaeological excavations.

There is some indication of the numbers of dogs that were used as food from the dance anklets, made of dog teeth, that survive in museum collections. Dog canine teeth were drilled for attachment to fabric bands and worn as dance rattles. One, in the Bishop Museum, has 11,218 teeth; it has been estimated that 2,805 dogs were required to produce the artifact.

Domestic dogs found their way from island to island across the Pacific Ocean. A major reason for their success as immigrants in the region was a total absence of competition from wolves, jackals, and foxes. Pacific Island canids themselves did not become feral, probably due mainly to a lack of adequate food in the wild, which meant they had to rely on their human owners for subsistence.

Today dogs are employed as food for human consumption in many parts of Asia, and China in particular, but the origins and reasons for this practice are not well documented, or if they are, they are probably recorded in one or more of the many Asian sources not yet translated.

Dog Fossils and Their Interpretation

Dog consumption practices in the Pacific and Asia notwithstanding, the dog's main attraction to humans, historically, has been its ability to hunt other animals selected as food by humans rather than serving as a source of food itself.

Over the past 50 years that I have been examining animal bones from archaeological sites, I have yet to see the bones of dogs bearing cut or butcher marks that would indicate that dogs had unquestionably been butchered for food. In the southwestern United States, for example, and particularly in Arizona, there are literally hundreds of articulated prehistoric dog

II.G.8 🐶 Dogs

The Dog as Human Food

The difficulty that confronts one at the very beginning of a study of the dog as human food is the lack of convincing evidence relating to the use of dogs (*Canis familiaris*) as food by early humans. Most of the discussions in published articles are either speculative or employ relatively recent evidence of dog consumption and extrapolate these data to the distant past.

There are some fine published discussions that relate to the eating of dogs in more recent times. One of the best is by Margaret Titcomb (1969), who documents the importance of the dog to human cultures in the ancient Pacific area. Titcomb believes (as do I) that the dog of the Pacific Islands was originally carried by canoe from the Asian mainland along with the pig and chicken for food on the voyage.

It was fed carefully upon vegetables such as taro and breadfruit so that its flesh was palatable to its

burials and many more isolated postcranial bones. The latter often show evidence of cutting, but it is evident that they were being prepared for bone artifacts. Most represent the shafts of humeri or femora and were scored by flint blades in transverse directions below the proximal articular condyle or the distal articular condyle, resulting in a bone tube when the articular ends were removed. Many of these artifacts, or "tubes," have been recovered from archaeological sites along with the discarded ends of the bones.

Some archaeologists have used the evidence of burned or charred dog bones to indicate meal preparation. However, experiments with fresh material have proven that burned or charred bone is the result of grilling to the point where edible flesh has been completely destroyed. Indeed, it must be destroyed before any change is apparent on the bones.

Generally, the sample size of dog burials is inadequate to determine their value to the associated human population. However, there are several occurrences of multiple dog burials that suggest reasons other than culinary for the buried animals. In Askelon, Persia, 1,238 dog skeletons of all ages, associated with a period dated between 822 and 428 B.C., were excavated and examined by archaeologists (Wapnish and Hesse 1993). There was no evidence of butchering or food preparation associated with these dogs.

Similarly, a site of puppy burials from the first-century-A.D. Roman villa at Lugnano, Italy, was excavated in June 1992. At least some of these dogs were associated with child burials, which may have had some ritual significance. In fact, dogs in association with human burials are not uncommon. Intentional burials of dogs, both separately and with humans, are known from Chinese sites (Figure II.G.8.1). Some of the dogs are in the same coffins with the humans; others have their own separate burial boxes.

Indeed, there is always a danger of misinterpreting dog remains when only bones are present. One of the best examples of this potential for misinterpretation is connected with a 1901 photograph taken on the shore of Penzhina Bay, in eastern Siberia on the Sea of Okhotsk. The photograph shows four large sled dogs hanging by their necks atop long poles that were set into the frozen earth (Perkins 1981). Waldemar Jochelson, the photographer (on what was an American Museum of Natural History expedition), explained:

All along the bank of the Paren River were stakes driven into the snow, with dogs hanging on them, their muzzles pointing upward. In the light of the spring sun, this long row of dog-sacrifices offered a queer and sad sight. I found out that the greater part of the dogs had been killed by the drivers of the sledges in gratitude for their safe return from Gizhiginsk, and to guard their villages from the measles, which a year previous had come to them from that little town. (Perkins 1981)

The dogs were eventually cut down from the poles and buried, and there was no evidence left indicating the circumstances that led to their deaths. It is doubtful that any excavators of the dogs would be able to glean the method of their execution or the reason for their demise. The point is that ritual sacrifices of dogs are not generally recorded. One example is a fenced brick-lined pit, off to one side of the market in Kathmandu in Nepal, whose purpose is for the sacrifice and dismembering of dogs. If it were not for the tourist photographs taken when the rituals are in progress, the small fenced pit might be mistaken for a dry fountain.

At sites in Egypt and prehistoric southwestern North America, dog mummies have been recovered. In the former region, humans did the mummifying, whereas natural causes were responsible in the latter. From these discoveries we can derive information about dog size and color of pelage and can make comparisons with surviving breeds of dogs. Two of the best-preserved animals are a small black-and-white dog and a yellowish-white one (Figure II.G.8.2) from a Basketmaker site at Marsh Pass, Arizona.

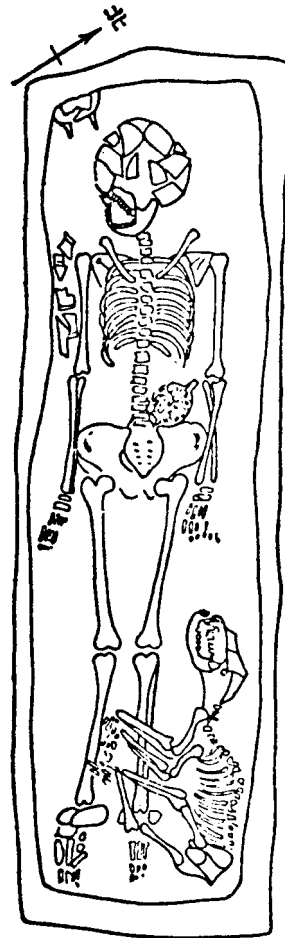


Figure II.G.8.1. Intentional burials of domestic dogs and humans. Neolithic of China; Xiawanggang c. 4000 B.P. (Illustration courtesy of Qi Guo-gin.)

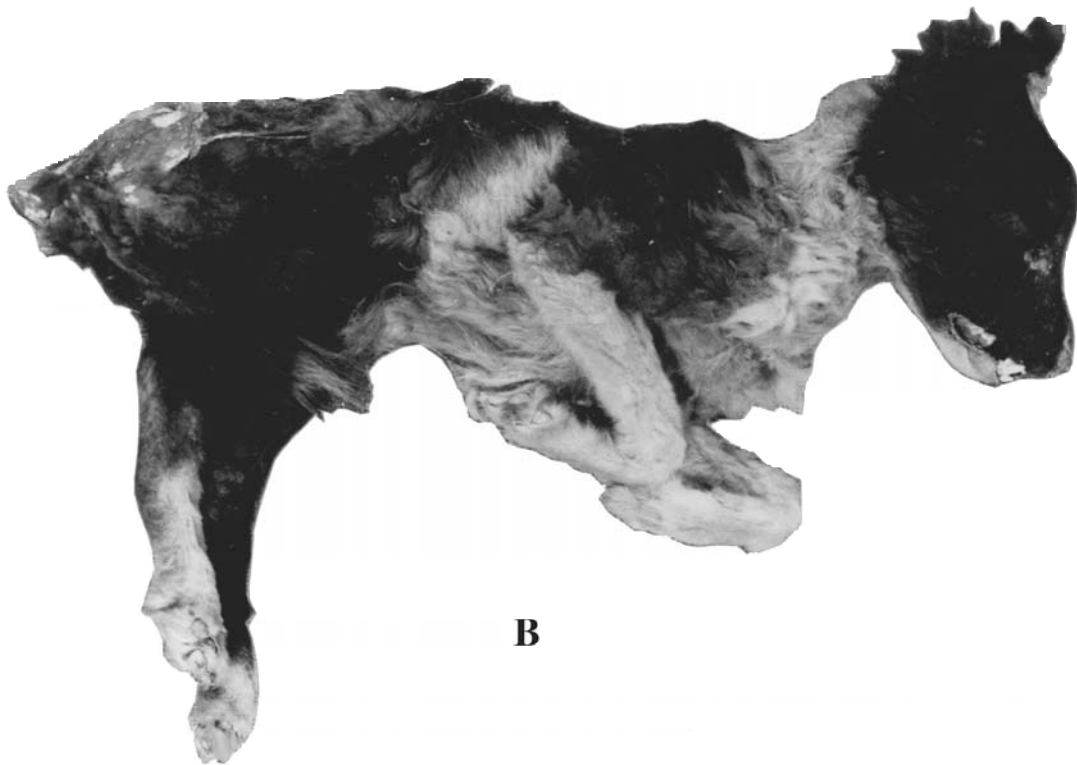
**A****B**

Figure II.G.8.2. Indian dogs, 2,000 years old, early Basketmaker. Natural mummies from White Dog Cave, Marsh Pass, Arizona: (A) yellow-white domestic dog; (B) black-and-white domestic dog. Both are *Canis familiaris*. (Photograph by John W. Olsen.)

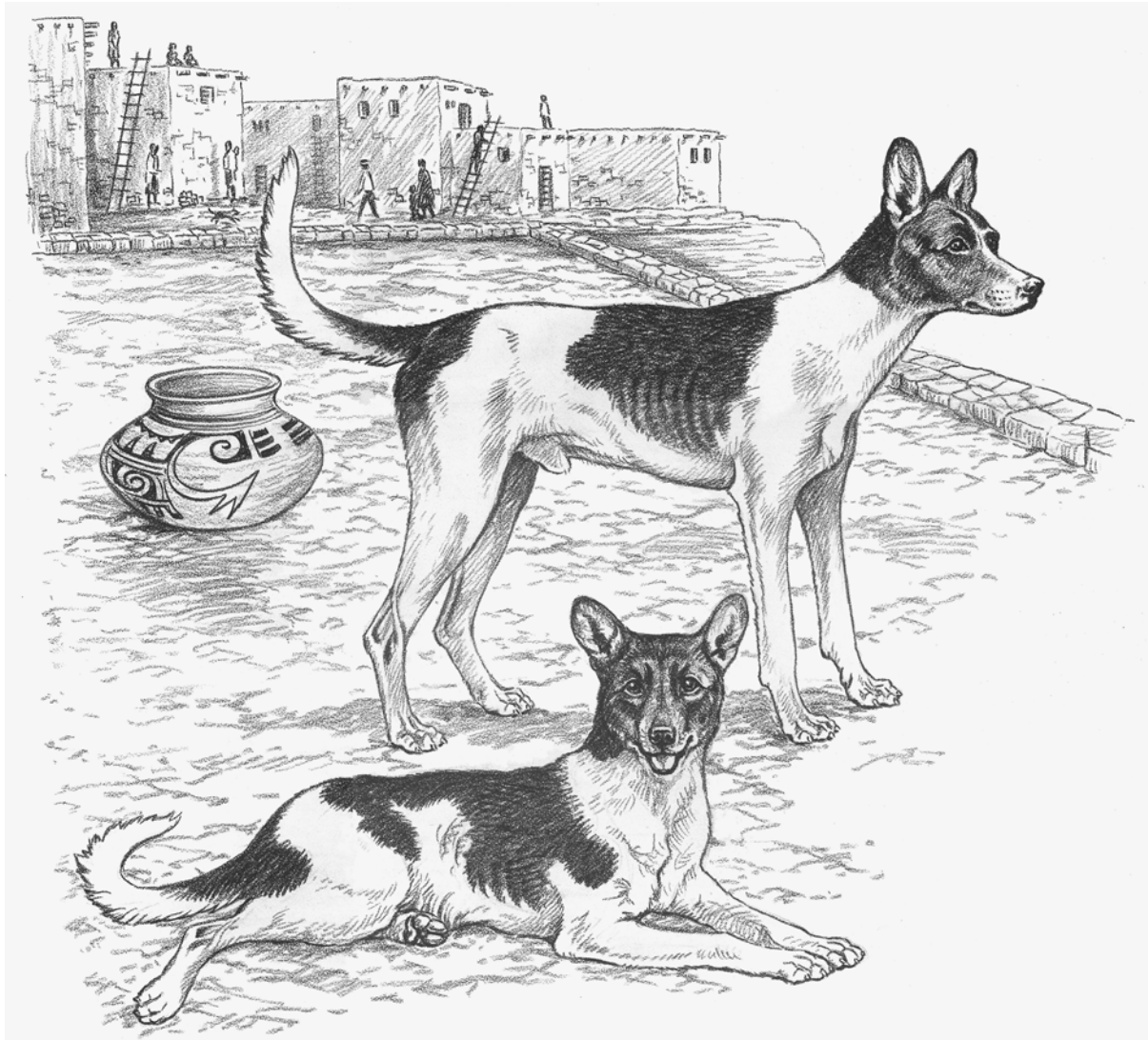


Figure II.G.8.3. Typical long-legged Basketmaker domestic dogs, *Canis familiaris*, from vicinity of Marsh Pass, Arizona, 2,000 B.P. (Drawing by Wallace Hughes.)

These dogs were mummified naturally in the hot, dry environment and placed in White Dog Cave with a mummified Pueblo Indian wrapped in a woven shroud. The dogs were a bit longer legged than the mongrels found in the vicinity today, but the color patterning and small size are rather common in the general area of Marsh Pass (Figure II.G.8.3).

Dog Domestication

If the pages of the past are blurred with respect to the use of the dog as food, they are equally hard to read in matters of dog domestication, which constitutes the burden of most of the remainder of this chapter. Much of what has been and is being pub-

lished on the events leading to canid taming and domestication is based more on supposition than on hard evidence. Woven into what follows is my own version of this process (based on Olsen 1985).

Studies relating to the social structures of both humans and wolves shed some light on what *may* have taken place in the initial development and relationship of these two groups of early hunters that eventually culminated in the close relationship between dogs and humans prevailing today throughout much of the world. Foremost among these studies are those by L. D. Mech (1970), M. W. Fox (1971), and B. H. López (1978). Perhaps the best compilation that relates to this relationship is a series of papers edited by R. L. Hall and H. S. Sharp (1978). Several theories

dealing with the attitudes of humans toward wolves and of wolves toward humans are discussed in detail in their volume.

Human hunter-gatherer societies and wolf packs were similar in a number of respects. Both comprised social units that were relatively small in number. Both were capable of hunting over open ground or wooded areas, pursuing rather large game and exerting considerable physical effort and energy over prolonged periods to accomplish their goals. Both used hunting methods that required a pack or team effort as opposed to the tactics of a lone hunter.

Faunal evidence from paleoanthropological sites of the late Pleistocene age (c. 10,000 B.C.) indicates that *Homo sapiens*, at this stage of development, depended on a diet that included a sizable portion of game animals. As carnivores, the wolves of this period also depended on wild game, and because hunting game by running it down takes a good deal more time than consuming the kill, one of the characteristics of both these predator groups would have been sharing the spoils of the hunt. Such socialization would have developed independently among both wolves and humans (Hall and Sharp 1978).

Another social feature shared by wolves and humans would have been that gender did not necessarily determine who did the hunting. Rather, such a determination would have depended on the individual's maturity and ability to hunt successfully. The nonhunters of the pack (immature animals), therefore, would gather in rendezvous areas that were selected for regrouping and for the sharing of freshly killed game provided by the active hunting members of the pack (Mech 1970; Hall and Sharp 1978; López 1978). The wolf packs' methods of obtaining game appear to be well organized and planned and should be briefly abstracted here as a possible model of how human hunting societies may have obtained game during the Pleistocene.

As pack hunters, wolves are able to capture and kill animals much larger than themselves. Artiodactyls such as deer, mountain sheep, elk, caribou, and even the moose, a formidable adversary, are all chosen prey. Rabbits, beaver, and (at times) fish in shallow water, as well as other small mammals, are also caught and eaten (Mech 1970). Wolves have been observed to change their hunting tactics more to suit the terrain and weather conditions than to suit the habits of a particular animal that they are pursuing. Their ability to procure food is not always successful; rather, feast or famine may be more the norm for these animals.

The selected prey in an ungulate herd may be an old or weakened animal, but this is not necessarily the case. Healthy, prime adults, as well as immature animals, are also selected and killed by wolves (Mech 1970). The method of bringing down such an animal is for the wolf to run alongside, slashing and tearing at its hindquarters, flanks, and abdomen until it is weak enough from its injuries to be dispatched (López

1978). Wolves may grab an animal, such as the moose or caribou, by the nose, holding on, while other members of the pack close in for the kill (Mech 1970; Hall and Sharp 1978; López 1978).

It is the case, however, that wolves generally choose animals that are easier to procure than healthy, active prey. Animals that are injured, infected by parasites to such a degree that they are weakened, or foundering in deep snow are all known to have been selected by wolf packs for food (López 1978). Animals in this condition, of course, signal their plight to some degree; these signals would not be lost on keen observers, such as wolves (or humans), and the weakened prey would be relatively easy to dispatch.

If a prey animal runs, it is certain to be pursued, especially if it is a lone individual (Mech 1970). To weaken such an animal, particularly if it appears to the wolf to be a strong and healthy adult, the practice of relay hunting may be put into practice, taking advantage of the terrain to wind the pursued. The hunt may end in a rush, terminating the animal's struggles in a few moments, or the dogged pursuit may go on for miles before the pack closes in. Alternatively, a wolf pack of four to six animals may send out two wolves to herd a lone victim into an ambush of the remaining members of the group (Mech 1970).

Wolves that have maintained a territory for a considerable length of time may know the terrain well enough to use shortcuts to intercept running prey. They are also known to hunt into the wind when approaching a feeding herd of artiodactyls.

The similar pack-hunting methods and social structures of early humans and wolves would not, in themselves, have been responsible for the first taming and domestication of the wolf. Perhaps humans observed the wolf's ability to track successfully by scent, or its ability to cover ground swiftly enough to elude human pursuit. It is possible that hungry wolves were enticed (not necessarily by human design) to come close to a campfire, where meat was being cooked and the refuse discarded in the immediate vicinity of the camp. Perhaps wolves that had attached themselves loosely to human habitation areas would consider such camps as their home territory, and their warning growls toward intruders would also warn the human inhabitants of the approach of such outsiders. If such an association occurred, it is not unreasonable to assume that these events would bring about firmer ties between wolves and humans to their mutual advantage. Unfortunately, evidence of the kind needed to prove or disprove this speculation is not the sort that could be preserved in the archaeological record.

If a hunter were to have initiated a closer association with wolves by taking a wolf pup from its den (possibly if a pup's parents were killed), this action would have encouraged imprinting of the substitute human family leader on the young wolf's behavior

pattern as it developed and matured in close association with the human hunting group. Thus, J. P. Scott (1950) wrote:

Behavior studies seem to indicate that the reason for the apparently easy and successful domestication of the wolf lies in the fact that it is a species with highly developed patterns of social behavior and that a large number of these patterns are sufficiently similar to human ones to permit mutual social adjustment between man and wolf. The primary stimuli are so similar in the two species that appropriate and recognizable social behavior is evoked.

But it is also quite possible that there existed a mutual respect between the wolf and human of 10,000 to 15,000 years ago for the hunting prowess of the other. At such a time, humans were themselves probably as rugged and aggressive as the wolves with which they shared a hunting territory.

In order for any animal, including wolves, to be successfully tamed and domesticated, it must be able to suppress its natural or normal living pattern and subjugate it to that which is dictated to it by humans. The wolf (in most instances), and the resulting domestic dog, have been able to do this, as have all of the common domestic animals that are familiar to us. But some animals cannot cross this bridge from wild to domestic. Records extending as far back as early dynastic Egypt indicate that many animals that seemed potential domesticates did not fulfill that potential. Some Egyptians experimented with hyenas, gazelles, and foxes, but they had little success in molding them into household animals. Similarly, the cape hunting dog (not a dog in the true sense of the word), the coyote, the jackal, and all of the large, wild felids were never brought beyond the tame or semi-tame stage. The modern wolf, however, appears to be an animal that will easily and fully accept the companionship of humans in a manner that nearly approaches that of a domesticated dog.

Since we are discussing the possibilities of the first friendly contacts between wolves and humans, it is appropriate to mention at this point the evidence we have for early domestication through osteological changes in the wolf. It is generally known and accepted by workers concerned with early canid domestication that the first observable changes take place in the skulls, jaws, and dentition of these tamed canids. More specifically, these alterations include the foreshortening of the muzzle, or rostrum, and the crowding of the tooth rows, along with a comparative overall reduction in the size of the teeth.

As domestication occurred, the mandibles deepened midway along the horizontal ramus, with a more convex inferior margin than that found in similar-sized wild wolves. These are all characteristics found in most domestic dogs of wolf size. The coronal apex takes a decided backward turn, a condition found in

some Chinese wolves (*Canis lupus chanco*) but generally absent in wolves from other geographic areas (Olsen and Olsen 1977). The foreshortening of the muzzle, or rostral, area of the skull has been observed in the remains of wolves recovered from geographically widely separated paleontological and archaeological sites of late Pleistocene age and later.

One of the earliest documentations of this foreshortening of wolf skulls was made by J. G. Pidoplichko (1969), who described short-faced, and possibly tamed, wolves. Their fossils were collected from excavations conducted in 1954 at a late Paleolithic campsite of mammoth hunters at Mezin in the Chernigov region of the Ukraine. They were found in association with a lithic assemblage amid the remains of 116 individuals of woolly mammoths (*Mammuthus primigenius*). The remains of wolves with proportions of those living in that area today (*Canis lupus albus*) were also found, along with the three or four short-faced animals.

Pidoplichko coined the taxonomic designation *Canis lupus domesticus* for these short-faced wolves. I, however, would propose to change the taxonomic assignment to *Canis lupus familiaris* for these and other similarly differing wolves and to use the name designated for the domestic dog to identify this Pleistocene race or subspecies, rather than create a new trinomial designation. The use of *familiaris* follows the correct taxonomic progression from a wild wolf, through taming, to the domestic dog. This term seems to be less confusing than the name *domesticus*, which has commonly been used as the scientific name for the domestic cat (*Felis domesticus*, now *Felis catus*).

For a number of years, particularly during the 1930s, the Frick Laboratory of the American Museum of Natural History in New York City had field representatives stationed in the Fairbanks area of Alaska. They were there to cooperate with the University of Alaska and the Fairbanks Exploration Company in collecting faunal remains as they were uncovered in muck deposits by hydraulic gold-mining operations. Such placer-mining methods, however, made it impossible to determine the stratigraphic context of a particular bone or artifact, because many had been collected by the gold-dredge crews from the spoil dumps of reworked and discarded matrix.

In fact, there is a general consensus that artifacts from such areas, which may originally have been on the surface, were subsequently relocated several meters below the ground level and mixed with the bones of extinct animals (Rainey 1939). Thus, it might have been possible to find some artifacts in association with the older, extinct animals if different methods of mining had been practiced when the bulk of known Pleistocene vertebrate material was collected in the 1930s.

Between the years 1932 and 1953, 28 more or less complete wolf skulls were obtained from the muck

deposits in an area north and west of Fairbanks. The age of the deposits from which these wolf skulls were obtained has been determined as Upper Pleistocene, Wisconsin Age (about 10,000 B.C.). The geologist T. L. Péwé published two reports (1975a, 1975b) on the geology and stratigraphy of this area and the associated recovered fauna. The wolf skulls were collected from dredging operations on Ester Creek, Cripple Creek, Engineer Creek, and Little Eldorado Creek. All of these creeks also yielded Paleolithic artifacts (Rainey 1939). However, the association of these artifacts with the skeletal remains of extinct animals is problematical, because the mechanical mixing of artifacts and bones, just discussed, prevented the collectors from obtaining any evidence that might have existed of a contemporary association of Pleistocene wolves with humans.

T. Galusha, of the Department of Vertebrate Paleontology at the American Museum of Natural History in New York City, worked for many years on the wolf remains from the Fairbanks dredging operations. He noted that a number of the skulls were extremely short-faced for wild wolves and approached modern Eskimo dogs in facial proportions, although they were considerably larger animals overall.

He turned the project over to me shortly before his death, and I continued his studies and compared this collection with a large series of both Pleistocene and Holocene wolves as well as with Eskimo dogs from both Greenland and Siberia. I feel confident at this point in stating that the short-faced wolves from the Fairbanks area appear to be forerunners of the later, domesticated Eskimo dogs.

The proportions of the skulls of these wolves that vary do so in the rostral area. The area of the skull that is anterior to the infraorbital foramen is noticeably foreshortened and constricted laterally in several of the skulls. Two of them (F:AM70933 and F:AM67156) lack anterior premolars. Several others have the full complement of teeth in the abbreviated dental margin. The dentition of F:AM67153 is considerably smaller than the average wolf of that area and approaches the overall tooth size found in the Eskimo dog.

Dishing of the rostrum, when viewed laterally, is evident in all of the short-faced skulls identified as *Canis lupus* from the Fairbanks gold fields. The occipital and supraoccipital crests are noticeably diminished compared to those found in average specimens of *C. lupus*. The occipital overhang of these crests, a wolf characteristic, is about equal in both groups of *C. lupus*. Multivariate analysis of all of the wolves from the Fairbanks collection separated the collection into two groups, based on either their predominant wolf-like or dog-like characteristics. Yet, as has already been noted, it may never be possible to establish if there was a close human association with the Alaskan Pleistocene wolves, because of the types of deposits in which both the wolves and the lithic assemblages

occur and the conditions under which both were collected.

In 1939 and 1940 an Eskimo village site was uncovered that dated from the first millennium B.C. This village, named Ipiutak, is believed to represent an occupation of early Eskimo immigrants from Siberia. The artifacts suggest a Neolithic culture (Larsen and Rainey 1948). The site is located on the tip of Point Hope, Alaska, approximately 125 miles north of the Arctic Circle, and is at the westernmost point of the continent north of the Bering Strait.

At least five more or less complete *Canis* skulls were recovered from the archaeological excavations at Ipiutak. These were studied and reported on by O. J. Murie in an appendix to the site report by H. Larsen and E. Rainey (1948). After comparing these skulls to those of wolves, Eskimo dogs, and other large domestic dogs, Murie decided that they fell into the Siberian class of dogs. Any slight differences between the Ipiutak dog skulls and the Siberian and Alaskan dog skulls with which they were compared were within the size variation considered to be normal for domesticated dogs of this type.

One animal in particular (No. H43) was considered to be a dog-wolf hybrid, as the large size and heavier mandible and teeth approached those of a wolf rather than a dog. Murie concluded that two of the skulls represented the so-called Siberian husky type, three may have been variants of this same type, and the other animal (No. H43) was the hybrid just mentioned. He believed that there was adequate evidence that the Ipiutak dogs were of Asian origin.

Both sides of the Bering Strait have seen numerous explorations and excavations searching for the earliest evidence of habitation sites of the Asian migrants who reached the North American continent. Unfortunately, very little relating to vertebrate remains has been reported. This may be partly because of the poor preservational conditions that characterize this region and that have resulted in an extremely small sample upon which to base our interpretations. However, the subordinate role of organic remains in many early archaeological interpretations has also had a great deal to do with this gap in our knowledge of fossil vertebrates of all taxa from archaeological contexts.

An example of this subordinate role can be seen in the 331-page report on the archaeology of Cape Denbigh (Giddings 1964). The text and 73 plates are devoted to a thorough coverage of nearly all aspects of the excavations on Cape Denbigh, but the recovered faunal remains are listed in a table of only seven lines in one paragraph. No scientific taxonomy is given for these animals, and some remains are classified simply as "bird" or "other." Obviously, it is quite possible to overlook many comparatively small canid fragments if one's interests are funneled in other directions. It is also possible that canids were not collected because it may have been assumed that they represented the common, local wolves, when in actu-

ality they might have represented primitive dogs, as well as wolves or crosses between wolves and dogs.

An intriguing collection of seven large canid skulls that may represent both domestic dogs and wolves or even wolf-dog crosses were found at the Bagnell site, which is a two-component village situated on the edge of a terrace above the Missouri River flood plain just north of Sanger, North Dakota. The dates of the proveniences from which the skulls were collected are from as early as circa A.D. 1590, with the latest occupation from the earliest layer yielding two skulls of domestic dogs of wolf size. Two more of the skulls are definitely those of wolves, but from the later time of occupation; one skull is of a domestic dog, also of wolf size, and the remaining two are of wolves. A most intriguing observation is that one of the domestic dog skulls from the early layer possesses two strong wolf characteristics and seven morphological characteristics that are typical of large, domestic dogs. Admittedly there is a problem of temporal control for these proveniences; however, at this time it does not seem likely that the earliest proveniences contain any intrusive mixtures of dogs that were introduced by Europeans. Large, wolf-sized dogs, particularly those that appear to be wolf-dog hybrids, are so rare in an archaeological context that these Bagnell canids are worthy of considerable note.

Another published account of a large, hybrid cross of a wolf and a dog is that contained in the report of excavations of a bison-kill site in southwestern New Mexico (Speth and Parry 1980). This find, a single canid skull, is illustrated and analyzed in the report, but unfortunately no provenience is given for this specimen. The overall range of dates for all proveniences on the site are from as early as A.D. 1420 $\pm$ 125 to A.D. 1845 $\pm$ 100. This canid would be of interest in relation to early taming of the wolf and, perhaps, its crossing with smaller Indian dogs, but only if it could be established that it was collected from a provenience that predated the European introduction of large mastiffs or hounds. Since the range of dates is so great, it could easily be the result of a more recent crossing of a European dog and local wolf. (The occurrence is recorded here as a plea to future workers to include all pertinent data and not just that which pertains to comparative zoology.)

To date, the oldest substantiated finds of prehistoric domestic dogs in North America are those from Jaguar Cave, Idaho. Fossil fragments, consisting of incomplete mandibles, a single left mandible, and a small portion of a left maxilla, were all collected from excavations in an early Holocene rock-shelter in the Beaverhead Mountains of Lemhi County, Idaho. The excavations were conducted by a joint expedition of the Peabody Museum of Archaeology and Ethnology at Harvard University and the Idaho State Museum at the Idaho State University in 1961 and 1962. Carbon-14 (^{14}C) dates indicate that the age of the deposits ranges from about 9500 to 8400 B.C. (Lawrence

1967). The site was determined to be a hunting camp (Sadek-Kooros 1972).

An interesting aspect that remains unexplained concerning the canid from this site is its unwolflike appearance (Lawrence 1967), which suggests there may be an even earlier form, as yet undiscovered, that may link these Jaguar Cave dogs with a wild, ancestral form. One would not expect to find these early dogs in a locality so far south as the Jaguar Cave rock-shelter without finding similar remains in sites closer to the Bering Strait. The morphological features that are present in the Jaguar Cave dog fragments are characteristically the same as those that are present in known specimens of *C. familiaris*. For example, critical measurements taken of all fragments indicate that they are too small to be derived from the wolf. They were then compared to the coyote, but they differ from this smaller, wild canid in being more massive, deeper dorsoventrally, and thicker lateromedially. The tooth rows are short when compared to the size of the individual teeth, and this shortening of the jaws is accompanied by crowding of the tooth row. This aspect is particularly noticeable in the area of anterior premolars, where the alveoli do not lie in a straight line; they are, rather, set obliquely. The muzzle, as far as could be projected from the fragments, is more shortened than in the coyote, and this characteristic seems to be well developed, as in later domestic dogs.

Some lower jaws of what are surely coyotes also occurred in the same site. These were determined to be of the local subspecies *Canis latrans lestes*. Canid material from a considerably later site (also in Birch Creek Valley, Idaho), having a ^{14}C date not earlier than 2500 B.C., was also described by B. Lawrence (1967) as belonging to domestic dogs. The specimens included four nearly complete mandibles, a number of mandibular fragments, one broken skull, and two cranial fragments. The specimens from this later site were determined to be very similar in form and size to Eskimo dogs, with which they were compared in the Museum of Comparative Zoology at Harvard University. By inference, the specimens from Jaguar Cave were also from this group of dogs.

There exists a considerable temporal gap between the Jaguar Cave dogs and later prehistoric dogs from other areas in North America. Fossils from areas in the Southwest date from at least the time of Christ (Guernsey and Kidder 1921). From other areas they date perhaps from 8400 B.C. (McMillan 1970). A dog from the Koster site in Illinois (Struever and Holton 1979) has an assigned date of 6500 B.C., and dogs from White's Mound in Richmond County, Georgia, have been given a date of about 500 B.C. (Olsen 1970). F. C. Hill (1972) briefly noted a Middle Archaic dog burial from Illinois with an accompanying date of 5000 B.C.

G. M. Allen (1920) published the first comprehensive discussion of early dogs that were associated with the prehistoric peoples of the Western Hemisphere.

This publication, unfortunately, is now out of date as well as out of print. Most of the critical finds relating to the development of the domestic dog were made subsequent to Allen's work. W. G. Haag (1948) based his evaluations of an osteometric analysis of some eastern prehistoric dogs on Allen's early publication, but Haag's monograph is also out of date.

The monograph by Allen was for many years the standard work that was used to classify domestic dogs found in association with prehistoric human cultures. The author listed these dogs as falling into the following groups, determined by comparative skull measurements and form: "(1) a large, broad-muzzled Eskimo dog, (2) a larger and (3) a smaller Indian dog, from which are probably to be derived several distinct local breeds. Of the larger style of dog, as many as eleven varieties may be distinguished; of the smaller, five" (Allen 1920: 503).

Some years ago I placed southwestern prehistoric dogs into two assemblages, small and large, following, more or less, Allen's classification. The animals in the small group were fox terrier-sized and were further refined or separated into small, short-faced, and small, long-faced dogs. The animals in the large group were long-faced and were comparable in size to the local coyote but a bit heavier in overall proportions. This latter group was referred to by Allen as the large Pueblo Indian dog, or the Plains Indian dog.

However, after examining later published reports and newer finds and reexamining the older finds made during Allen's career, I now conclude that these groupings are, in a sense, artificial – particularly for dogs from the southwestern United States. In fact, the groups actually grade into one another in size, form, and amount of morphological variation, if a large enough collection is examined. The result is a more or less single mongrel group of southwestern Indian dogs, although the entire range of size and form variation is not found in every archaeological site. It is still possible to find representatives of only a part of the spectrum – either small, short-faced or long-faced dogs, or large Pueblo Indian dogs – at specified sites. The overlap is at the extremes of each of these size groups and is quite logical, because the prehistoric Indian dogs were hardly registered American Kennel Club breeds but were, instead, free-breeding, socializing mongrels.

It is, of course, quite simple to pick out representative animals of these differing groups from collections of excavated canids if there are enough individuals assembled from a large number of archaeological sites. The Basketmaker and early Pueblo Indian dogs of the Southwest and those from White's Mound, Georgia, as well as the dogs from the shell heaps of Kentucky or Alabama, all show a close similarity in size and form. But these animals were quite advanced domestic dogs, and most were of a comparatively small size, although – as stated earlier – there are exceptions to this rule (for example, both small and

large forms of dogs are found in the same stratigraphic level at the Jaguar Cave site).

Many early and later Pueblo Indian sites have yielded dog remains. In particular, reference is made to immature puppies as well as adults found in levels dating from about A.D. 1125 to 1175 at Antelope House in Canyon de Chelly, Arizona (Kelley 1975). At the Grasshopper Pueblo, built during the fourteenth century in east central Arizona, a number of small domestic dogs were recovered during 16 years of excavation by the University of Arizona Archaeological Field School (Olsen 1980). A unique find of an immature gray wolf, *C. lupus*, consisting of a right premaxilla and deciduous dentition, was also recovered from one of the rooms of this pueblo. This fragment of a wolf pup suggests that it may have been kept as a pet, perhaps with a view toward taming, although this is only speculation.

It is the case, however, that nearly every Pueblo excavation in the southwestern United States has produced some evidence of the domestication of the dog. These sites date from the eleventh through the fourteenth centuries and range in size from Cliff Palace Ruin in Mesa Verde National Park, Colorado (with its multistoried masonry construction of many rooms and great stone towers), to the more modest pueblos at Keetseel and Betatakin in the area of Monument Valley, Arizona. Tree-ring dates for Keetseel range between A.D. 1274 and A.D. 1284; those for Betatakin are between A.D. 1260 and A.D. 1277. By the time of these late dates, the domestic dogs were well advanced and were morphologically the same as present-day dogs.

As of this writing, however, we have yet to recover the ancestral forms of *C. familiaris* that would help to bridge the gap between the small, late Paleolithic wolves and the early Neolithic dogs from Asia and from Jaguar Cave, Idaho, on the one hand, and, on the other hand, the gap between the Jaguar Cave dogs and the well-known series of dogs of the southwestern United States that date from some 2,000 years ago and are still in existence today.

Stanley J. Olsen

Bibliography

- Allen, G. M. 1920. Dogs of the American Aborigines. *Bulletin of the Museum of Comparative Zoology, Harvard University* 63: 431-517.
- Fox, M. W. 1971. *Behavior of wolves, dogs, and related canids*. New York.
- Giddings, J. L. 1964. *The Archaeology of Cape Denbigh*. Providence, R.I.
- Guernsey, S. J. and A. V. Kidder. 1921. Basketmaker caves of northeastern Arizona. Monograph in *Papers of the Peabody Museum of Archaeology and Ethnology*, Vol. 8, No. 2. Cambridge, Mass.
- Haag, W. G. 1948. An osteometric analysis of some aboriginal

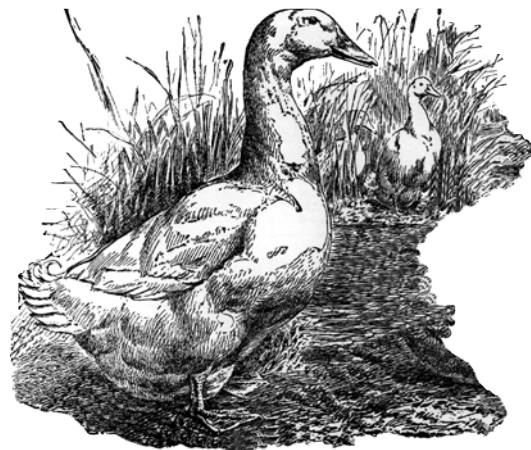
- dogs. Monograph in *Reports in Anthropology*, Vol. 7, No. 3. Lexington, Ky.
- Hall, R. L., and H. S. Sharp. 1978. *Wolf and man*. New York.
- Hill, F. C. 1972. *A Middle Archaic dog burial in Illinois*. Evanston, Ill.
- Kelley, J. E. 1975. Zooarchaeological analysis at Antelope House: Behavioral inferences from distributional data. *The Kiva* 41: 81-5.
- Larsen, H. and F. Rainey. 1948. Ipiutak and Arctic whale hunting culture. Monograph in *Papers of the American Museum of Natural History*, Vol. 41. New York.
- Lawrence, B. 1967. Early domestic dogs. *Sonderdruck aus S. F. Säugetierkunde* 32: 44-59.
- López, B. H. 1978. *Of wolves and men*. New York.
- McMillan, R. B. 1970. Early canid burial from Western Ozark Highland. *Science* 167: 1246-7.
- Mech, L. D. 1970. *The wolf*. Garden City, NY.
- Olsen, J. W. 1980. A zooarchaeological analysis of vertebrate faunal remains from Grasshopper Pueblo, Arizona. Ph.D. dissertation, University of Michigan.
- Olsen, S. J. 1970. Two pre-Columbian dog burials from Georgia. *Bulletin of the Georgia Academy of Science* 28: 69-72.
1985. *Origins of the domestic dog*. Tucson, Ariz.
- Olsen, S. J., and J. W. Olsen. 1977. The Chinese wolf ancestor of New World dogs. *Science* 197: 533-5.
- Perkins, J. 1981. *To the ends of the earth*. New York.
- Péwé, T. L. 1975a. *Quaternary geology of Alaska*. Geological Survey Professional Paper 835. Washington, D.C.
- 1975b. *Quaternary stratigraphic nomenclature in unglaciated Central Alaska*. Geological Survey Professional Paper 862. Washington, D.C.
- Pidoplichko, J. G. 1969. *Late Paleolithic dwellings of mammoth bones in the Ukraine*. Kiev.
- Rainey, F. G. 1939. Archaeology in Central Alaska. *Anthropology Papers of the American Museum of Natural History* 36: 351-405.
- Sadek-Kooros, H. 1972. Primitive bone fracturing: A method of research. *American Antiquity* 34: 369-80.
- Scott, J. P. 1950. The social behavior of dogs and wolves: An illustration of socio-biological systematics. *Annals of the New York Academy of Sciences* 51: 1001-122.
- Speth, J. D., and W. J. Parry. 1980. *Late prehistoric bison procurement in southeastern New Mexico. The 1978 season at the Garnsey Site (LA-18399)*. Ann Arbor, Mich.
- Struever, S. and F. A. Holton. 1979. *Koster: Americans in search of their prehistoric past*. Garden City, NY.
- Titcomb, M. 1969. *Dog and man in the ancient Pacific with special attention to Hawaii*. Honolulu, Hawaii.
- Wapnish, P., and B. Hesse. 1993. Pampered pooches or plain Pariahs? The Ashkelon dog burials. *Biblical Archaeologist* 56: 55-79.

by humans and probably domesticated in different areas at different times. The amount of variability in the domestic duck is very small compared with that found in the domestic chicken (Thomson 1964), and it would seem that present-day domestic ducks evolved gradually (Woelfle 1967), in the process becoming larger than the wild type, with much more variety in color, size, and gait (Clayton 1984). Domestic ducks have also lost the ability to fly.

The excellent flavor of duck flesh (as well as the eggs) has been enjoyed from prehistoric times to the present day. An important incentive in breeding ducks for meat has been the fact that they have a fast growth rate and can be killed as young as 6 to 7 weeks of age and still be palatable. A disadvantage, however, is that duck carcasses are very fatty (Clayton 1984).

Ducks are raised in large numbers in many Western countries, such as the Netherlands, Britain, and the United States, although intensive duck production has occurred only in the last 20 years (Clayton 1984). In Britain, most commercial ducks are found in Norfolk (although some are kept in Aberdeen and Dumfries), but these constitute only about 1 percent of all poultry in the country (Urquhart 1983). Ducks are less prone to disease than hens but eat more food. Unfortunately, their eggs are unpopular with British consumers because they are thought to be unclean. Ducks destined for the supermarkets are killed when they are from 7 to 9 weeks old.

It is in the East where fully 75 percent of domestic ducks are found, especially in the Asian countries of Vietnam, Indonesia, Thailand, China, Bangladesh, the Philippines, and Burma. Duck meat contains a high percentage (up to 35 percent) of fat, which is in short supply in many Asian diets, and in China, where less meat is consumed on a regular basis than in the Western world, selection is for ducks with a high fat content (Clayton 1984).



Peking Duck

II.G.9 🦆 Ducks

The mallard, *Anas platyrhynchos*, is the most ubiquitous taxon in the subfamily *Anatinae* of the family *Anatidae*. It is the ancestor of most domestic ducks, the males of which still sport the ancestral curling feathers of the upper tail (Delacour 1956-64; Gooders 1975; Gooders and Boyer 1986). Because the wild mallard is so widespread in the Northern Hemisphere, it is extremely likely that it was widely utilized

Present-day domestic ducks range from the small Call, weighing less than 1 kilogram (kg), to the larger strains of Peking and Aylesbury, weighing up to 6 kg. Ducks can be grouped into three categories: meat producers (Rouen and Aylesbury), egg layers (Indian Runner and Khaki Campbell), and ornamental birds (the tiny Call and Crested ducks) (Delacour 1956–64). The Normandy or Rouen duck is quite similar to the wild mallard and, although common worldwide, is of greatest economic importance in Southeast Asia. The bird has a gamy flavor and in France is traditionally killed by strangulation or smothering, thus ensuring that the flesh retains all of its blood and flavor (David 1970).

Most domestic breeds have a characteristic horizontal posture, but some (as, for example, the Indian Runner), are much more erect, which allows them to run very quickly. The selective pressures that encouraged this erectness are not known, nor are the anatomical changes that brought it about, although droving, common for centuries in the Far East, may be a possibility (Clayton 1972, 1984).

Ducks in the Prehistoric Period

The earliest work on bird bones from archaeological sites was carried out in the mid-nineteenth century at shell mounds in Jutland, Denmark (Steenstrup 1855). Little attention, however, was given to the identification of bird bones from archaeological sites in the Northern Hemisphere until the 1980s, and scant research has been undertaken on the taphonomy of bird bones (Ericson 1987; Livingston 1989). Another problem is that morphological differences between similar bones of different taxa of the *Anatidae* are not always distinct and clear-cut. Measurements of bones can help to separate them, but frequently no specific identification can be made, especially if the material is fragmentary. Nevertheless, attempts at distinguishing wild and domestic mallard bones have been made, although there are also zooarchaeological problems in such an undertaking. Anatomical differences that may be associated with domestication have been recognized in several late Saxon faunal collections in Britain, including that of Northampton (Coy 1981, 1989). The flattened facet on the anterior aspect of the *caput femoris*, employed by J. Lepiksaar (1969) for the identification of domestic geese, can also be used for ducks. J. Ekman (1973) applied this technique to remains in seventeenth- or eighteenth-century deposits in Gothenburg, Sweden.

The wild mallard is (and was) highly migratory, with flocks of several thousand birds not unusual (Gooders 1975). Ideally, wild birds are hunted in the late summer and early fall, just after they have grown plump on summer feeding, and prehistoric fowlers generally took the birds at this, their most defenseless period, or while they were molting or nesting in the spring (Clark 1948).

Some Danish Mesolithic wetland sites demonstrate

human exploitation aimed almost solely at birds and, in certain cases, seasonal exploitation of just one or a few species. For example, the Mesolithic site of Aggersund on the Limfjord was a specialized camp for the procurement of whooper swans (Grigson 1989).

There are two distinct types of Danish Mesolithic sites. One is the inland bog, associated with the Maglemosian culture of the Boreal age, whereas the other is the coastal midden of the Ertebølle culture of the Atlantic age. The chief Maglemosian sites on Zealand (Mullerup, Svaerdborg, Holmegaard, and Øgaarde) have yielded bird bone assemblages in which the most abundantly represented birds are wild ducks and mute swans. Also represented are grebes, coots, and some sea birds, including cormorants, gulls, and divers of various species (Clark 1948). Marine birds are more strongly represented in the Ertebølle middens. Because of a lack of winter migrants, the Maglemosian sites were probably occupied only during the summer months, but fruitfully so, as evidenced by the presence of cranes and numerous mute swans, as well as young cormorants and sea eagles.

The mute swan and mallard are both resident taxa and remain very common in present-day Denmark (Grigson 1989). Mallards breed all over inland Denmark and molt during the summer in most wetland habitats, usually in small flocks, but occasionally several hundred individuals congregate on large lakes and in densely vegetated marshland.

In addition to the work done in Denmark, we know from efforts in Germany (at Friesack, Landkreis Nauen) that the Mesolithic (9,700 to 7,000 years ago) bird bone assemblage is dominated by the mallard in all phases (Teichert 1993). In Neolithic eastern Europe (western Russia, the northeastern Baltic region, and northern Poland), numerous bird bones have been excavated from wetland settlements. It is thought that some of these may have been specialized waterfowling sites (Zvelebil 1987). Eighty-nine percent of the bone sample at Narva-Riigikula 3 is composed of waterfowl, mainly ducks. The site is situated on a sandy dune separating a lagoon from the open sea and is on the route of the annual migration of waterfowl between their breeding grounds in the Arctic and their winter quarters in southern Scandinavia and northwestern Europe. These wetland sites are of immense importance in preserving bone and plant remains as well as food-processing tool kits – in this case, wooden projectiles with blunted tips – thus making it possible to reconstruct the subsistence strategies of forest-zone hunter-gatherers in a much more comprehensive way than at dry sites.

As was the case in Poland, the wild duck was the chief quarry of the Iron Age inhabitants of Glastonbury Lake Village in Somerset, England. There, the birds were probably dispatched by clay pellets (also excavated from the site) that were well adapted for use with a sling (Andrews 1917).

Ducks in the Historical Period

The Evidence of Domestication

Southeast Asia is claimed to have been a major center of duck domestication (Zeuner 1963; Wood-Gush 1964), especially southern China, where the birds were kept during the Earlier Han Dynasty (206 B.C. to A.D. 220). The first written records of domestic ducks date back to the Warring States period (475–221 B.C.) (Weishu 1989). But according to one authority, the Chinese have had domesticated ducks for at least 3,000 years (Yeh 1980, cited in Clayton 1984), and it is the case that Chinese pottery models of ducks and geese, dating from about 2500 B.C., have been excavated (Watson 1969). However, archaeological evidence of a faunal nature is also needed before any firm conclusions about the antiquity of domestic ducks in China can be reached.

As already noted, Southeast Asia continues today as an important duck-raising area where domestic ducks surpass chickens in economic importance (Zeuner 1963). The duck thrives in watery environments, which promote clean plumage and increased disease resistance. Moreover, in warm, humid climates, with abundant rice paddies and waterways, ducks forage more successfully and produce more eggs than do chickens, which are not fond of this sort of environment (Clayton 1984).

In China, ducks are particularly important in the control of land crabs (which can devastate rice crops); the birds eat the crab nymphs that eat rice seedlings. Ducks are also released into paddy fields to consume locust nymphs, which, unlike adult locusts, are unable to fly. The recognition that ducks are effective instruments in combatting insect pests is probably of great antiquity (Needham 1986).

The main duck taxa winter in southern China and, except during the breeding season, move in large flocks, each consisting of hundreds – or even thousands – of birds. Although many different species of domestic duck have evolved in Jiangsu and other provinces south of the Yangtze River, where duck raising is common, their shapes, colors, and wings still bear much resemblance to those of their ancestors. The well-known Peking duck, for example, evolved through domestication of the mallard. Its feathers have whitened over a long period, but the curling central tail feather shows that the bird is descended from the same ancestor as the domestic ducks in the south (Weishu 1989).

True domestication elsewhere, however, seems to have come much later. Surveying the limited textual evidence (from Aristotle in the fourth century B.C. up to the ninth century A.D.), J. Harper concluded that ducks did not become fully domesticated outside of China until possibly the Middle Ages (Harper 1972).

The pintail (*Anas acuta*) is the most frequently represented taxon of waterfowl in ancient Egyptian art and hieroglyphics (Houlihan 1986), and such illus-

trations include depictions of pintails being force-fed. These birds can be tamed easily and remain able to breed (Delacour 1956–64). The widespread marshes of the Nile Delta would have provided excellent wintering grounds for both ducks and geese and, therefore, fowling opportunities for humans. But although geese, chickens, and pigeons are frequently mentioned in Egyptian papyri from the fifth century B.C. onward, ducks are noted only rarely.

Aristotle discussed only chickens and geese in his *Natural History*, and although Theophrastus mentioned tame ducks, he failed to indicate whether they were bred in captivity (Harper 1972). F. E. Zeuner (1963) has asserted that the keeping of domestic ducks in Greek and Roman times was unusual, though not unknown. Duck-shaped vases have been recovered at Rhodes and Cyprus, both centers of the cult of Aphrodite; they were dedicated to the goddess and to her companion, Eros. Ducks may have been on these islands for religious purposes only.

Several species were kept in captivity by the Romans, who maintained aviaries (*nessotrophia*) of wild ducks, probably to fatten them for the table (Toynbee 1973). Varro, writing in 37 B.C., was the first to mention duck raising by the Romans, pointing out that ducks should be enclosed to protect them from eagles as well as to prevent their escape (Hooper and Ash 1935). In the first century A.D., Lucius Junius Moderatus Columella provided advice on keeping ducks (mallard, teal, and pochard) and other waterfowl in captivity, which was considered much more difficult than caring for more traditional domestic fowl (Forster and Heffner 1954).

Columella, along with Marcus Tullius Cicero and Pliny the Elder, recommended that the eggs of wild ducks be gathered and placed under hens to hatch. Columella claimed that ducks raised in this way would breed readily in captivity, whereas birds taken as adults would be slow to commence laying. He also stated that duck keeping was more expensive than the raising of geese, which fed mainly on grass (Forster and Heffner 1954). Marcus Porcius Cato mentioned the fattening of hens, geese, and squabs for the market, but not ducks (Hooper and Ash 1935).

Moving from Rome to other parts of Europe, we find that metrical data from Colchester, England (from the first to the fourth centuries A.D.), indicate that length and breadth measurements of mallard wing and leg bones are within the range of variation for wild mallards (Luff unpublished). Similarly, measurements of mallard remains from the Roman fort of Velsen I in the Netherlands were considered to be of the wild variety; length and width measurements of the mallard bones fit perfectly within the ranges of recent wild mallards studied by H. Reichstein and H. Pieper (1986; also see Prummel 1987).

The Saxons may have had domestic ducks, but as yet the evidence is unclear. Remains unearthed at Saxon Hamwih (Southampton) could have been

those of domesticated ducks but could also have been those of wild birds (Bourdillon and Coy 1980). However, J. P. Coy (1989) has suggested that the low proportion of wildfowl at Saxon sites in Britain is indicative of the economic importance of domestic birds such as ducks and chickens.

A bit later, in Carolingian France (the eighth to the tenth centuries A.D.), estate surveys listing payments due feudal lords indicate that chickens and geese served as tender far more frequently than ducks (Harper 1972). Similarly, in Germany, the *Capitularium de Villis* grouped ducks with ornamental birds such as peafowl, pheasants, and partridges, but not with chickens and geese, kept for purely economic reasons (Franz 1967).

The scarcity of wildfowl was most likely significant in hastening domestication. Early medieval England, for example, had apparently witnessed a wholesale slaughter of wildfowl in spite of Acts of Parliament specifically aimed at their conservation, and in 1209, King John, finding insufficient game for his personal falconry, issued a proclamation forbidding the taking of wildfowl by any means (Macpherson 1897). Statutes were also passed by Henry V, and again by Henry VIII, against the destruction of wildfowl. Interestingly, in the price controls of the Poulterers' Guild in London, dated to 1370, a distinction was drawn between wild and tame mallards (Jones 1965). Wild mallards were more expensive than tame ones, which suggests that wild birds were more prized than farmyard birds and indicates their general scarcity at that time.

Jean Delacour (1956-64) has suggested that the mallard may have become truly domesticated in Europe only in the medieval period - although, of course, prior to the distinction made by St. Hildegard (twelfth century) between wild (*silvestris*) and domestic (*domesticus*) ducks (Figure II.G.9.1). Certainly, the latter were being reared in France by the end of the fourteenth century, when the *Ménagier de Paris* (c. 1390) also distinguished between wild and domesticated ducks (Prummel 1983).

Nonetheless, ducks of domestic origin are uncommon in archaeological contexts of the medieval period (Eastham 1977; Maltby 1979). At early medieval Dorestad (Netherlands), some ducks were larger than others, indicating that they might well have been domesticates (Prummel 1983). However, the majority of remains at that site probably derived from wild specimens, because their measurements fall within the range for modern wild mallards (Woelfle 1967). Many mallard remains from late medieval Amsterdam measured even larger than those from Dorestad, once again suggesting the presence of domestic ducks among the Dutch (Prummel 1983). But by contrast, at Haithabu in northern Germany, duck remains have been firmly identified as deriving from the wild variety (Requate 1960; Reichstein 1974), which was also the case at eleventh- and

twelfth-century Grand-Besle, Buchy, Normandy (Lepiksaar 1966-8).

Although domestic ducks are often identified in archaeological deposits from the sixteenth century onward, they did not increase dramatically in size until the eighteenth and nineteenth centuries, when distinct varieties were recorded. Methods of rearing ducks were modeled on those mentioned by classical authors. G. Markham, for example, in the seventeenth century described ways of keeping wild mallards, teals, widgeons, shellducks, and lapwings that were very similar to those mentioned by Columella (Markham 1614).

In the eighteenth century, breeders began to promote and further develop certain traits for frequent egg laying or rapid growth for meat production, and sometimes breeds were crossed to produce hybrids suitable for both purposes (Batty 1985). Other varieties were used mainly as ornamental birds to decorate gardens and park ponds.

Aspects of Domestication

S. Bökönyi has emphasized the differences between "animal keeping" and "animal breeding," with the former occurring without purposeful selection or the control of feeding, whereas the latter involves the deliberate selecting of specific traits for animal breeding and also control of nutritional intake (Bökönyi 1969).

There are a number of possible reasons why duck domestication in the West lagged so far behind that of the goose and chicken. One is that goslings accept the first living creature that they see as their mother, whereas mallards, who imprint through sound, do not do this. For them it is the call note of the mother that is important in identifying her (Lorenz 1964). Consequently, ducks were much less amenable to domestication, and the Roman idea of placing wild duck eggs under hens to hatch them into domesticity was off the mark because it did not go far enough.

Interestingly, the Comte de Buffon (Georges Louis Leclerc), who authored 32 volumes of *Natural History*, explained how such a procedure did work:

Eggs taken from the reeds and rushes amidst the water, and set under an adopted mother, first produced, in our farm-yards, wild, shy, fugitive birds, perpetually roving and unsettled, and impatient to regain the abodes of liberty. These however after they had bred and reared their own young in the domestic asylum became attached to the spot and their descendants in process of time grew more and more gentle and tractable, till at last they appear to have relinquished and forgotten the prerogatives of the savage state, although they still retain a strong propensity to wander abroad. (quoted in Bewick 1826)

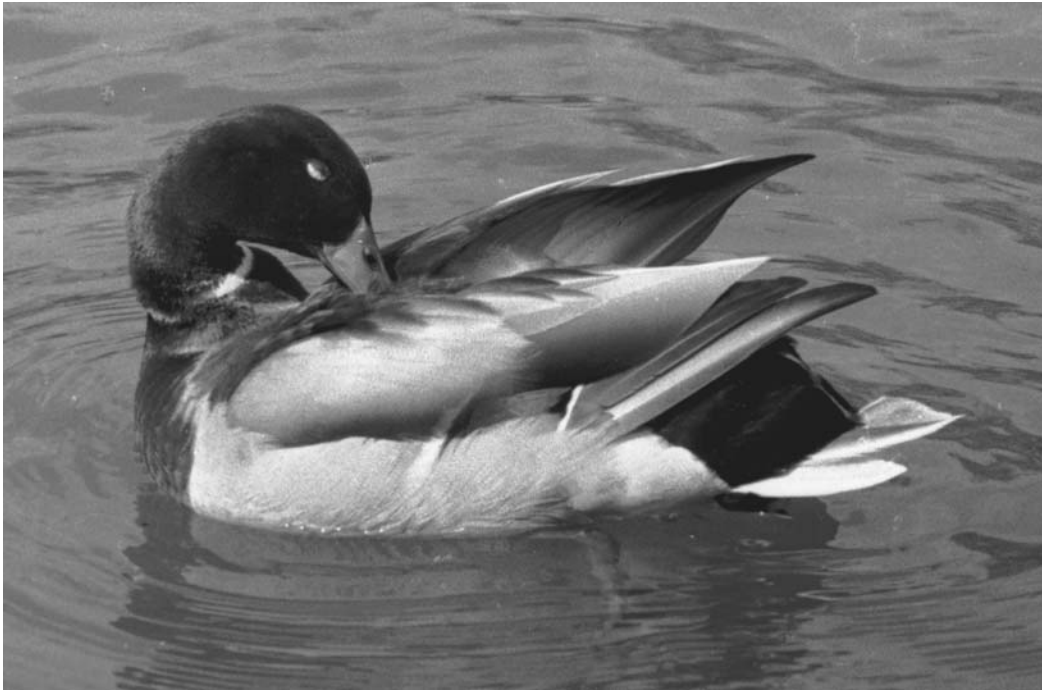


Figure II.G.9.1. The mallard duck, ancestor of most domestic ducks.

Another reason for what appears to have been tardy domestication in the West has to do with temperature. S. Bottema (1989) has pointed out that domestic fowl in the Near East always have larger numbers of young than those in more temperate regions, which is the result of a higher survival rate connected not with greater clutch size but with higher temperatures.

Temperature, in turn, is linked to diet. If eggs of the teal (*Anas crecca*) are hatched when temperatures are low, the young may not survive, even though teal ducklings in the wild are not affected by cold weather. The reason is that a natural diet in the wild compensates for a lack of body warmth, but in captivity, with a suboptimal diet, a warm temperature becomes a critical factor (Bottema 1989). In fact, it is for this reason that Bottema (1989) has proposed that areas with high spring temperatures (such as much of Asia) would have been the locus for initial duck domestication.

Charles Darwin (1875) was one of the first researchers to deal with morphological changes in domestication. Using a small sample of wild mallard, Aylesbury, Tufted, Penguin, and Call ducks, he found that: (1) In comparison with the wild duck, the domestic duck experienced universal but slight reduction in the length of the bones of the wing relative to those of the legs; and (2) the prominence of the crest of the sternum relative to its length was much reduced in all domestic breeds. In addition, E. Brown (1906: 5), citing Edward Hewitt's breeding of wild ducks in 1862, has commented that "the beautiful carriage of the wild mallard and his mate changes to the easy, well-to-do, comfortable deportment of a small Rouen, for they at each reproduction become much larger."

Yet although most researchers, such as Darwin and Brown, have equated domestication with a size increase, E. P. Allison has suggested that there was a size decrease in mute swans from the Neolithic to the Bronze Age in Cambridgeshire, and that the birds were, on average, significantly larger than recent specimens. This phenomenon is not related to climatic change (Northcote 1981, 1983). But it is related to confinement, which can produce smaller birds that have not been selected for intentionally, such as the shellduck (*Tadorna tadorna*) or the male pintail (*A. acuta*). It is also related to food conditions in captivity, which can bring about a size decrease (Bottema 1989).

Confinement and food conditions in captivity may also have affected ducks in ways we have yet to understand. Another confounding factor in faunal analysis of mallards from riverine urban sites is that they may have arisen from a number of different populations, including wild, domestic, and scavenging ducks, with the latter most likely to interbreed with the former (O'Connor 1993). Thus, it is debatable just how pure the strains of present-day river

mallards are, as they are obviously the product of considerable inbreeding. Ducks can hybridize (doubtless this was responsible for genetic contributions to domestic stock), and, in fact, all breeds of domestic duck can interbreed.

Also bearing on the history of duck domestication has been the birds' fussiness (much greater than that of chickens), which forestalls their development in caged coops because they have large appetites and because their feet are not tolerant of the wire floors. Ducks produce proportionately larger eggs than gallinaceous birds (and their young seem more resistant to starvation) (Marcström 1966), but selection under domestication has been for a larger bird and not for a larger egg. Moreover, duck eggs, as already noted, have never been in high demand in Europe; they are strong in flavor and lack the palatability of hens' eggs (Brown 1930). This is unfortunate, as ducks are naturally prolific layers and the eggs are highly nutritious.

Summary

Current evidence suggests that the origin of duck domestication was in Southeast Asia, and particularly in China. This region is still an important center for duck breeding and consumption, and one where captive, newly hatched ducklings have a greater chance of survival because of relatively higher temperatures.

It is true that more zooarchaeological data – in the form of bones – would be helpful in determining the origins of duck domestication. But because distinctions between the wild and domestic types of ducks are slight, such evidence alone would probably not pinpoint these origins.

Although the Egyptians, Greeks, and Romans kept ducks in captivity, they never truly domesticated them. Compelling evidence for domestication outside of Southeast Asia has only been found for the medieval period and later. In Europe, and particularly in Britain, the widespread pursuit and destruction of waterfowl by the nobility during the early Middle Ages (and probably earlier) may well have provided the impetus for duck domestication.

Because ducks have voracious appetites, feeding them may often have been a problem during winter months, when food shortages were common in non-industrial societies – especially after a poor harvest. Consequently, it is likely that even though tamed, ducks were frequently left to forage as best they could. This inference also suggests that breeding was not controlled, with tame ducks interbreeding with members of the wild population. It was only in the eighteenth and nineteenth centuries that specialized breeds of domestic ducks were reared for meat, eggs, and ornamentation.

Rosemary Luff

Bibliography

- Allison, E. P. 1985. Archaeozoological study of bird bones from seven sites in York. Ph.D. thesis, University of York.
- Andrews, C. W. 1917. Report on the remains of birds found at the Glastonbury Lake Village. In *The Glastonbury Lake Village*, by A. Bulleid and H. St. George Gray, Vol. 2, 631-7. Glastonbury, England.
- Batty, J. 1985. *Domesticated ducks and geese*. Liss, England.
- Bewick, T. 1826. *A history of British birds*, Vol. 2. London.
- Bökönyi, S. 1969. Archaeological problems and methods of recognizing animal domestication. In *The domestication and exploitation of plants and animals*, ed. P. J. Ucko and G. W. Dimbleby, 219-29. London.
- Bottema, S. 1980. Fenden in Nederland. In *Zelzame buisdier-rassen*, ed. A. T. Clason, 191-204. Zutphen Thieme, the Netherlands.
1989. Some observations on modern domestication processes. In *The walking larder*, ed. J. Clutton-Brock, 31-45. London.
- Bourdillon, J., and J. Coy. 1980. The animal bones. *Excavations at Melbourne St., Southampton, 1971-76. Council for British Archaeology Research Report* 33: 79-121.
- Brown, E. 1906. *Races of domestic poultry*. London.
1930. *British poultry husbandry, its evolution and history*. London.
- Clark, G. 1948. Fowling in prehistoric Europe. *Antiquity* 22: 116-30.
- Clayton, G. A. 1972. Effects of selection on reproduction in avian species. *Journal of Reproduction and Fertility* (Supplement 15): 1-21.
1984. Common duck. In *Evolution of domesticated animals*, ed. I. L. Mason, 334-9. London.
- Coy, J. P. 1981. The bird bones. In *Excavations in Chalk Lane, Northampton*, by J. H. Williams and M. Shaw. *Northampton Archaeology* 16: 1-9.
1989. The provision of fowls and fish for towns. In *Diet and crafts in towns*, ed. D. Serjeantson. *British Archaeological Reports* 199: 25-40.
- Darwin, Charles R. 1875. *Variation of animals and plants under domestication*. London.
- David, E. 1970. *French provincial cooking*. Harmondsworth, England.
- Delacour, Jean. 1956-64. *The waterfowl of the world*, Vols. 2, 3, and 4. London.
- Eastham, A. 1977. Birds. In *Excavations at Portchester Castle*, ed. B. Cunliffe. Reports of the Research Committee of the Society of Antiquaries of London 34: 233-9.
- Ekman, J. 1973. Early medieval Lund, the fauna and the landscape. In *Archaeologica lundensia. Investigationes de antiquitatibus urbis Lundae* 5. Lund, Sweden.
- Ericson, P. G. P. 1987. Interpretations of archaeological bird remains: A taphonomic approach. *Journal of Archaeological Science* 14: 65-75.
- Forster, E. S., and E. H. Heffner. 1954. *Columella: De re rustica*, Vol 2. London.
- Franz, G. 1967. *Capitularium villis. Quellen zur Geschichte des deutschen Bauernstandes im Mittelalter*. Darmstadt, Germany.
- Gooders, J. 1975. *Birds, an illustrated survey of the bird families of the world*. London.
- Gooders, J., and T. Boyer. 1986. *Ducks of Britain and the Northern Hemisphere*. London.
- Grigson, C. 1989. Bird-foraging patterns in the Mesolithic. In *The Mesolithic in Europe*, ed. C. Bonsall, 60-72. Edinburgh.
- Harper, J. 1972. The tardy domestication of the duck. *Agricultural History* 46: 385-9.
- Hooper, W. D., and H. B. Ash. 1935. *Cato and Varro: On agriculture*. London.
- Houlihan, P. F. 1986. The birds of ancient Egypt. In *The natural history of Egypt*, Vol. 1. Warminster, England.
- Jones, P. E. 1965. *The worshipful company of poulterers of the city of London*. Oxford.
- Lepiksaar, J. 1966-8. *Restes d'animaux provenant du Grand Besle, Meddelanden från Lunds Universitets. Historiska Museum*, 85-116.
1969. Nytt om djur från det medeltida Ny Varberg. *Varbergs Museums Arsbok*, 1-32.
- Livingston, S. D. 1989. The taphonomic interpretation of avian skeletal part frequencies. *Journal of Archaeological Science* 16: 537-47.
- Lorenz, K. 1964. *King Solomon's ring*. London.
- Macpherson, H. A. 1897. *A history of fowling*. Edinburgh.
- Maltby, M. 1979. The animal bones from Exeter, 1971-75. *Exeter Archaeological Reports* 2: 66-74.
- Marcström, V. 1966. Mallard ducklings (*Anas platyrhynchos*) during the first days after hatching. *Viltrety* 4: 343-70.
- Markham, G. 1614. *Cheap and good husbandry*. London.
- Mason, I. L. 1984. *Evolution of domesticated animals*. London and New York.
- Mead, W. 1931. *The English medieval feast*. London.
- Needham, J. 1986. Biology and biological technology. In *Science and civilization in China*, eds. J. Needham and Lu Gwei-djen. Vol. 6, Part 1, Sec. 38. Cambridge.
- Northcote, M. 1981. Size differences between limb bones of recent and subfossil mute swans, *Cygnus olor*. *Journal of Archaeological Science* 8: 89-98.
1983. Morphology of mute swans in relation to domestication. In *Animals and archaeology*, Vol. 2., ed. J. Clutton-Brock and C. Grigson. *British Archaeological Reports, International Series* 183: 173-9.
- O'Connor, T. P. 1993. Birds and the scavenger niche. *Archaeofauna* 2: 155-62.
- Pichon, J. 1847. *Ménagier de Paris (le), traité de morale et d'économie domestique composé vers 1393 par un bourgeois parisien*. Paris.
- Prummel, W. 1983. *Excavations at Dorestad 2. Early medieval Dorestad, an archaeozoological study*. Rijksdienst voor het Oudheidkundig Bodemonderzoek, Amersfoort, the Netherlands.
1987. Poultry and fowling at the Roman castellum Velsen I. *Palaeohistoria* 29: 183-201.
- Reichstein, H. 1974. Ergebnisse und Probleme von Untersuchungen an Wildtieren aus Haithabu (Ausgrabung 1963-1964). In *Untersuchungen an Tierknochenfunden (1963-1964)*, ed. H. Reichstein and M. Tiessen. *Berichte über die Ausgrabungen in Haithabu* 7: 103-44. Neumünster, Germany.
- Reichstein, H., and H. Pieper. 1986. *Untersuchungen an Skelettresten von Vögeln aus Haithabu. (Ausgrabung 1966-1969)*. Neumünster, Germany.
- Requate, H. 1960. Das Hausgeflügel. In *Die Haustiere von Haithabu (Die Ausgrabungen in Haithabu, 3)*, ed. W. Herre, 136-46. Neumünster, Germany.
- Seeböhm, M. E. 1952. *The evolution of the English farm*. London.
- Steenstrup, J. 1855. Et bidrag til Gerrfuglens, Alca impennis Lin., naturhistorie, og saerligt til Kundskaaben om dens tidligere Udbredningskreds. *Videnskabelige Med-*

- delelser fra den naturhistorisk Forening Kjobenbaun 3: 33-116.
- Teichert, L. 1993. Die Vögelknochenfunde von Friesack, Landkreis Nauen, eine paläoökologische und faunengeschichtliche Auswertung. In *Festschrift für Hanns-Hermann Müller. Zeitschrift für Archäologie* 27: 17-28.
- Thomson, A. L. 1964. *A new dictionary of birds*. London.
- Toynbee, J. 1973. *Animals in Roman life and art*. London.
- Urquhart, J. 1983. *Animals on the farm, their history from the earliest times to the present day*. London and Sydney.
- Watson, W. 1969. Early domestication in China. In *The domestication and exploitation of plants and animals*, ed. P. J. Ucko and G. W. Dimbleby, 393-5. London.
- Weishu, X. U. 1989. *Birds in China*. Beijing.
- Willughby, F. 1678. *Ornithology*. London.
- Woelfle, E. 1967. Vergleichend morphologische Untersuchungen an Einzelknochen des postcranialen Skelettes in Mitteleuropa vorkommender Enten, Halbgänse und Säger. Ph.D. thesis, University of Munich.
- Wood-Gush, D. G. M. 1964. Domestication. In *A new dictionary of birds*, ed. A. L. Thomson, 215-18. London.
- Yeh, Hsiangkui. 1980. *Fossil ducks and Peking ducks*. China.
- Zeuner, F. E. 1963. *A history of domesticated animals*. London.
- Zvelebil, M. 1987. Wetland settlements of eastern Europe. In *European wetlands in prehistory*, ed. J. M. Coles and A. J. Lawson, 94-116. Oxford.

II.G.10 🦌 Game

Evolutionary Underpinnings

Game (defined here as the meat of wild mammals, birds, and terrestrial reptiles) has been an important component of the human diet since earliest times. A signal difference between the digestive system of humans and those of their closest primate relatives is the modification of the guts of the former for the processing of a diet containing substantial quantities of meat (Chivers 1992). Until the last few thousand years, all meat-producing animals were wild. The continuing cultural significance of hunting as an activity, and game as an element of the diet, must be seen in the light of the age-old importance of game animals in human biological and cultural evolution.

Although little doubt exists that early hominids consumed appreciable quantities of game, lively debate lingers over whether our remote ancestors were primarily hunters or scavengers (e.g., Shipman 1986; Binford 1988). Nevertheless, there is abundant evidence that by the time of the appearance of early modern humans - between 50,000 and 100,000 years ago - hunting was one of humankind's major activities. It may have been so for uncounted previous millennia.

A striking sexual division of subsistence labor characterizes most contemporary hunting and gathering peoples, with men hunting and women gathering (but

see later discussion). Typically, gathered foods more or less reliably provide the bulk of the calories in the diet, mainly as carbohydrates, whereas game (and/or fish) contributes the undependable and sometimes limiting protein fraction. It has been proposed (Isaac 1978; Isaac and Crader 1981) that this bifurcation in foraging tasks and products is ancient and is implicated in the coevolution of food sharing, tool use, paternal investment in offspring, and related features of the unique human constellation of social behaviors.

Game, particularly large game, is often shared widely, whereas gathered foods tend to be acquired and consumed within the family. Meat is usually more valued than vegetable foods, and hunting is typically a more prestigious activity than gathering (Kent 1989). These correlations, too, may be ancient. The ancestral associations of men with hunting, of hunting with high social status, and of game with generosity and display, are an important part of the background of the presence of game in the human diet.

Subsistence Hunting

Global Trends

Studies of the hunting and gathering societies have identified geographic regularities in the quantity of game in the diet of these foraging peoples for whom wild foods are the foundation of life (Lee 1968; Jochim 1981). Behind these regularities are climatically determined abundances and availabilities of game animals relative to other potential human foods.

Close to the North Pole, inedible (for humans) tundra grasses dominate the vegetation, whereas much of the faunal biomass is in grazing animals (caribou, reindeer) with large herds and highly palatable flesh. Because much of the circumpolar region is coastal, terrestrial game can often be supplemented by large sea mammals (seals and whales). Thus, most of the few high-latitude foraging peoples (Eskimos, native Siberians) live largely or almost exclusively on game; the only real alternative food is fish (and that solely for inhabitants of uncommonly well-favored seabords).

A good number of cool-temperate hunting and gathering peoples (40°-60° latitude) have fish as the primary element in the diet, particularly anadromous fish (fish that run upriver from the ocean to spawn, frequently in enormous numbers; see Schalk 1977). These peoples, however, also consume considerable game. Well-known ethnographic examples are the inhabitants of the northwest coast of North America and the coastal islands of northeast Asia, locations endowed with rivers supporting reliable and abundant fish runs. Other cool-temperate foraging peoples (that is, the Cree and Ojibwa of eastern Canada and several bison-hunting tribes of the North American Great Plains) make game the basis of their diet and eat distinctly lesser quantities of fish and vegetable foods. Cool-temperate regions are mainly forested or covered with prairie grasses, and wild vegetable foods

are seasonally scarce and often not abundant at the best of times. With the exception of the North American bison, most of the large game tends to be relatively solitary and the availability of fish and game in a locality usually decides the base of the diet.

Hunting and gathering peoples living within 40° of the equator – the majority of all foraging peoples, and the great majority of those for whom we have reliable dietary information – derive most of their diet from gathered vegetable foods. Tubers, nuts, and seeds tend to be abundant and reliable relative to fish and game at these latitudes. Nevertheless, game remains for these peoples the most desirable part of the diet. Its capture attracts more effort and attention than the harvest of vegetable foods, and its procurers are often rewarded with favorable attention ranging from subtle deference to sexual favors (Hawkes 1990).

It is frequently assumed that game's desirability to these lower-latitude foragers derives from its role as the only major source of limiting nutrients, particularly the eight essential amino acids found in animal protein. However, complete inventories and nutritional analyses of these foragers' gathered wild foods remain scarce. That game and fish do supply much (usually most) of their protein is clear; that alternative invertebrate and vegetable protein foods are actually lacking is less clear. Well-documented demonstrations that gathered foods could not supply a protein-adequate diet are still too few to establish the generality of the impression that game is strictly necessary to provide protein for low-latitude hunters and gatherers.

Cultural Evolutionary Trends

Within the last 10,000 years (perhaps a bit earlier in Southeast Asia) plants and animals were domesticated on all continents except Australia. In some regions (e.g., eastern North America, lowland South America), the major edible domesticates were all plants. Here, wild game and fish continued to supply the meat in everyone's diet, not only in early agricultural times but even after domestication had permitted the emergence of stratified societies. In these New World stratified societies, the mighty were differentiated from the lowly in diet as in other aspects of life. However, it was the amount of game, the species of game, and the cuts of meat that distinguished the larders of the haves from those of the have-nots; there was no significant substitution of domestic flesh for wild game.

In other regions (that is, Europe, western Asia), large herd animals were domesticated at around the same time(s) as plants, and livestock became important early in the history of food production. Nevertheless, even though domesticated animals and their products (milk, blood) quickly became major elements in the diet of these mixed farmer-herders, game remained in the diet. The hunt continued in parallel with stock raising throughout most of early village life in the Old World, with wild animals often important

in the diet. As food production led eventually to the development of Old World complex societies, hunting survived. At that point, however, the reasons for its persistence – previously a straightforward matter of obtaining food – became divergent.

By the early days of stratified societies, a pattern had been established (e.g., Trigger 1965) that continued through much of history and into the modern world: Hunting, and the presence of game in the diet, was most prominent in the lives of people at the very top and the very bottom of the social hierarchy. Among rulers, hunting lingered as recreation and ostensive display, a demonstration of not needing to rely on the efficient meat production of domesticated animals. Among the peripheral and the poor, hunting persisted because people could not afford domestic animals. The conflict of these classes over access to game lies behind such beloved folklore as the legend of Robin Hood. Among both classes, and into modern times, hunting is pleasurable, and game is a prestige food, although there are higher prestige foods among the impoverished.

Sexual Division of Labor

Although numerous societies have produced the occasional female hunter, and many cultures recognize goddesses of the hunt (the Grecian Artemis or the Indian Chandi), hunting remains predominantly (often exclusively) a male pursuit in almost all societies. Reasons proposed for masculine dominion of the chase range from the advantage in mean strength and speed enjoyed by male hunters to the disadvantages in combining hunting with pregnancy and child care incurred by female hunters. In this connection, the evidence of the tropical forest Agta people of the Philippines is crucial: Among some bands of the hunting, gathering, casually horticultural, and trading Agta, women regularly hunt, and their hunting includes dangerous game – like the wild pig – that they kill close up with machetes and improvised spears.

It is significant, then, that among the Agta, most women's hunting involves dogs; that "during late pregnancy and for the first few months of nursing, a woman will not hunt" (Estioko-Griffin and Griffin 1981: 131); and that "[y]oung mothers . . . are not the major women hunters; mature women with self-sufficient children secure most of the game animals considered [of those taken by women] both by number and by weight" (Griffin and Griffin 1992: 310). Although the sexual division of labor appears remarkably weak among the Agta, and the generalization that women are mainly confined to gathering is overturned by this case, hunting, nonetheless, remains more associated with men than with women.

Optimal Foraging Theory

A body of ecological theory known as Optimal Foraging Theory or OFT (Stephens and Krebs 1986) generates predictions as to the prey species that should be

included in the diet of a predator. The general rule is that the predator should choose to pursue individuals of a particular species only when the predator cannot expect to get more or better food for its hunting time by disregarding that species. If there is a statistically reliable expectation that members of more valuable prey species will be taken in the time that would otherwise have been dedicated to the pursuit of members of a less valuable species, then that less valuable species is not in the optimal diet.

For example, if hares in a grassland are difficult to catch, but gazelles are plump and easy to kill, the optimal forager ignores the small and time-costly hares, even when he sees them bounding away, as long as gazelles are plentiful. Time spent chasing a hare is time taken away from searching for gazelles. Hares enter the optimal diet only when gazelles have become so rare that it becomes cost-effective to chase hares. When the mean encounter rate with individuals of each potential game species is known, along with the average time it takes to pursue and kill them and the average amount of meat they provide, then the optimal diet set can be predicted with considerable precision.

Predictions can also be made as to the predator's order of preference for the animals included in the optimal diet set. Predicted choices are easily summarized: Hunters should favor large animals over small and easily killed animals over those that require much time to hunt down and often escape.

Optimal diet set predictions have been tested for a number of subsistence hunters, with observed hunting behavior generally agreeing with the OFT model (Hill and Hawkes 1983; Kuchikura 1988; Kaplan and Hill 1992). Preferred game orderings have also been examined from an OFT perspective. Again, subsistence hunters appear in general to conform to theoretical predictions (Altman 1987; Hawkes, O'Connell, and Blurton-Jones 1991).

Game Allocation

The goal of subsistence hunting is game to eat. However, it is seldom if ever the case that all animals are considered game or that all parts of all game animals are eaten randomly by all members of a society. Rules governing the allocation of game are highly variable cross-culturally. Indeed, the issue of who makes the distribution is far from uniform from one culture to another. In many egalitarian cultures, the game belongs to the man (or woman) who killed it; in others, to the hunter who first wounded it; and in still others, to the person who made the arrow or other weapon with which the kill was accomplished, independent of that person's participation in the hunt. In ranked or stratified societies, the right to distribute the game may belong to the head of the lineage to which the successful hunter belongs, or to a village chief, or to another official.

Among the egalitarian, hunting and gathering

!Kung Bushmen of southern Africa, the hunter whose arrow first hit the animal makes the initial distribution of meat, which is to the other hunters in his party. These hunters, in turn, distribute shares to their kin. All the hunters' kin who receive meat have additional kin obligations, and eventually everyone in the camp receives some meat, although portions may be small for individuals genealogically distant from the members of the hunting party (Marshall 1965).

But distribution methods aside, there are often rules of striking salience as to where different parts of animals are directed. Among the egalitarian, horticultural Anggor of New Guinea, when a wild pig is killed, the carcass, once cooked, is parceled out in a detailed way: The extremities are given to local clans; the hunter's clan gets the viscera; old women eat the fetus of a pregnant sow; old men receive cooked blood; and so on (Hurber 1980).

Parts (or all) of particular animals may be reserved for or prohibited to particular people or all people. Among the egalitarian, horticultural Barí of Colombia and Venezuela, the head and tripe of howler monkey are not eaten by anyone, whereas the head of spider monkey is eaten, along with the liver and some other organ meats. Peccary heads are reserved for old men; children are not allowed to consume any part of a bear; and no one eats jaguar.

People in particular statuses may have privileges or prohibitions connected with consumption of designated parts of game animals. Excavations at sites belonging to the late prehistoric Mississippian chiefdom centered at Moundville, Alabama, indicate that preferred cuts of deer meat were sent from outlying areas to elite centers, presumably for consumption by chiefs, their families, and retainers (Welch 1991).

Entire game species may be associated with social statuses. Among contemporary Anglo-Americans, consumption of opossum meat is restricted to rural lower classes, although venison is acceptable at both extremes of the social ladder, and smaller game birds are eaten largely by people of means.

There are a few general principles known for game distribution: (1) Many cultures assign carnivores and scavengers to an "inedible" category; (2) in egalitarian societies, the larger and more unpredictably obtained a game animal is, the more likely it is to be shared widely; (3) in ranked and stratified societies, the largest and most desirable game animals and the best cuts typically go to the most important members of society; and (4) the reproductive parts of otherwise edible animals are often prohibited to some if not all members of a society. Beyond these empirical generalizations, a general theory of food allocations, prescriptions, and proscriptions remains to be constructed.

Game in the Diet

The amount of game in the victuals of subsistence hunting peoples varies from a small fraction, as among many peoples with domestic food animals

who also hunt when the opportunity arises (many societies in Oceania are of this type), to virtually the entire diet, as among some Eskimo groups. Some regional trends were mentioned previously. Within subsistence hunting societies, the distribution of game may or may not also be highly variable, with some or all game animals widely distributed and consumed or effectively confined to a minority of the society's members. In general, the higher the proportion of game in the aggregate diet, the weaker the exclusion of classes of individuals within the society from game consumption.

Prestige, Recreational, and Commercial Hunting

Prestige Hunting

Rulers of antiquity were often conspicuous hunters. Near the dawn of recorded history, "the Middle Assyrian king Tiglath-pileser I claimed . . . to have killed 4 wild bulls, 10 elephants and 920 lions – 800 from his chariot and 120 on foot" (Roaf 1990: 154).

Millennia later, in December A.D. 799, on a different continent, Charlemagne and his three sons killed a wolf, 2 European bison, 6 aurochs, 66 boars, and 80 deer in the space of two hours, to the applause of the queen and princesses, at a hunt organized by a count and his foresters. The royal party feasted later that day on the choicest parts of the kill (Halleux 1988).

Marco Polo's description, nearly 500 years later still, of Kublai Khan's yearly round illustrates a number of the features of prestige hunting: its prominence in royal life; its ostensive display; its employment of sumptuary laws; and its ranking of game animals by their suitability as royal prey. For six months of the year, hunting figured large in the life of Kublai Khan's court:

The three months of December, January, and February, during which the Emperor resides at his Capital City, are assigned for hunting and fowling, to the extent of some 40 days' journey round the city; and it is ordained that the larger game taken be sent to the Court. To be more particular: of all the larger beasts of the chase, such as boars, stags, lions, bears, etc., the greater part of what is taken has to be sent, and feathered game likewise. . . .

After he has stopped at his capital city those three months . . . he starts off on the 1st day of March, and travels eastward towards the Ocean Sea, a journey of two days. He takes with him full 10,000 falconers and some 500 gerfalcons besides peregrines, sakers, and other hawks in great numbers; and goshawks also to fly at the water fowl. . . . The Emperor himself is carried upon four elephants in a fine chamber made of timber, lined inside with plates of beaten gold, and outside with lion's skins. . . . He always keeps beside him a dozen of his choicest gerfalcons,

and is attended by several of his Barons, who ride on horseback alongside. . . . And when he has travelled till he reaches a place called Katar Modun, there he finds his tents pitched . . . full 10,000 tents in all. . . . The Lord remains encamped there until the spring [May] and all that time he does nothing but go hawking round and among the canebrakes . . . and across fine plains . . . for 20 days journey round the spot nobody is allowed . . . to keep hawks or hounds . . . and furthermore throughout all the emperor's territories, nobody however audacious dares to hunt . . . hare, stag, buck and roe, from . . . March to . . . October. (Polo 1927: 127, 129–33)

Recreational Hunting

Much prestige hunting is clearly recreational, but hunting for enjoyment is not limited to royalty. Indeed, most hunting, even explicitly commercial hunting, has a pleasurable aspect. Although the recreational chase of the upper classes is nowadays often more a matter of public socializing or trophy skins than of table meat (British fox hunting, international big cat hunting), a notable proportion of the recreational hunting of all classes produces edible game such as duck and deer. In modern times, ethical considerations may impel the consumption of this game as a justification for the enjoyment of the hunt, even when the hunter finds it less palatable than the highly processed food to which he is accustomed. Generations of North American boys have learned a version of the "You shoot it, you eat it" hunter's code from their fathers. The pleasures of the hunt may paradoxically cause moderns to eat game they would otherwise omit from their diets.

Commercial Hunting

Hunting in order to sell or barter game has roots more ancient than commonly recognized. There are no contemporary tropical forest hunters and gatherers who do not maintain trading relations – most often game for crops – with agricultural neighbors. It has been proposed that humans did not enter the tropical forest until the invention of agriculture made this interchange possible (Headland 1987; Headland and Bailey 1991). The Pygmies of the Congo Basin have been trading game for crops with their Bantu neighbors for centuries, if not millennia. Similar relations, possibly of comparable antiquity, are maintained with agriculturalists by the hunting and gathering Makú of the Amazon region, the Agta of the Philippines, and a number of Hill Tribes in India. The game-producing hunters and gatherers are typically considered by their agricultural trading partners to be serfs or slaves.

In Western industrial societies commercial meat hunters are also marginal people. Guides and "white hunters" are glamorous figures in fiction and may, in

fact, make considerable money, but they provide recreational and prestige hunting for others. They do not make a living by selling the game they hunt themselves. In a sense, contemporary individuals who sell game can be looked at as the cultural descendants of the peripheral individuals who continued to hunt for food while more affluent members of their societies obtained animal protein from domesticated animals.

Game is usually tougher than the meat of domesticated animals, but many consider it to be tastier. Its generally lower fat content has led some to believe that it is healthier as well.

It is indicative of the persistence of social evolutionary trends noted earlier that in the Western world, restaurants habitually serving wild game are of two kinds – the most expensive establishments in major metropolitan centers and small, “mom and pop” places in rural areas. Many people frequenting both kinds of establishments eat game in considerable measure for its status marker value, although of course the prestige of consuming game differs widely in association as well as scope between one type of restaurant and the other. In the former, game is affiliated with wealth and display, whereas in the latter it is primarily linked with recreation and the manly virtues.

Aside from its survival in this and similar commercial contexts, game has largely disappeared from the diet in modern industrial societies. Areas where it is still a notable part of yearly food consumption are uncommon. These include tribal enclaves, a few rural areas where subsistence hunting traditions are strong, and a very few spots where recreational and prestige hunting are major components of the local economy. In prospect, it seems that whatever amount of game remains in the diet of humans in the developed world will be retained for its association with the prestige of the hunt rather than for the nutritional value of the meat itself.

Stephen Beckerman

Bibliography

- Altman, J. 1987. *Hunter-gatherers today: An aboriginal economy in north Australia*. Canberra, Australia.
- Binford, L. R. 1988. The hunting hypothesis, archaeological methods, and the past. *Yearbook of Physical Anthropology* 30: 1-9.
- Chivers, D. J. 1992. Diets and guts. In *The Cambridge encyclopedia of human evolution*, ed. Steve Jones, Robert Martin, and David Pilbeam, 60-4. Cambridge.
- Estioko-Griffin, A., and P. B. Griffin. 1981. Woman the hunter: The Agta. In *Woman the gatherer*, ed. Frances Dahlberg, 121-51. New Haven.
- Griffin, P. B., and M. B. Griffin. 1992. Fathers and childcare among the Cagayan Agta. In *Father-child relations: Cultural and biosocial relations*, ed. Barry Hewlett, 297-320. Hawthorne, N.Y.
- Halleux, G. 1988. Charlemagne à la chasse. *Plaisirs de la chasse* 437: 28-30.
- Hawkes, K. 1990. Why do men hunt? Some benefits for risky choices. In *Risk and uncertainty in tribal and peasant economies*, ed. E. Cashdan, 145-66. Boulder, Colo.
- Hawkes, K., J. O'Connell, and N. Blurton-Jones. 1991. Hunting income patterns among the Hadza: Big game, common goods, foraging goals and the evolution of the human diet. *Philosophical Transactions of the Royal Society of London B* 334: 243-51.
- Headland, T. 1987. The wild yam question: How well could independent hunter-gatherers live in a tropical rain forest ecosystem? *Human Ecology* 15: 463-91.
- Headland, T., and R. Bailey. 1991. Introduction: Have hunter-gatherers ever lived in tropical rain forest independently of agriculture? *Human Ecology* 19: 115-22.
- Hill, K., and K. Hawkes. 1983. Neotropical hunting among the Aché of Eastern Paraguay. In *Adaptive responses of native Amazonians*, ed. Ray Hames and William Vick-ers, 223-67. New York.
- Hurber, P. 1980. The Anggor bowman: Ritual and society in Melanesia. *American Ethnologist* 7: 43-57.
- Isaac, G. L. 1978. The food sharing behavior of protohuman hominids. *Scientific American* 238: 90-108.
- Isaac, G. L., and D. C. Crader. 1981. To what extent were early hominids carnivorous? An archaeological perspective. In *Omnivorous primates: Gathering and hunting in human evolution*, ed. R. S. O. Harding and G. Teleki, 37-103. New York.
- Jochim, M. 1981. *Strategies for survival: Cultural behavior in an ecological context*. New York.
- Kaplan, H., and K. Hill. 1992. The evolutionary ecology of food acquisition. In *Evolutionary ecology and human behavior*, ed. Eric Smith and Bruce Winterhalder, 167-201. Hawthorne, N.Y.
- Kent, S. 1989. Cross-cultural conceptions of farmers as hunters and the value of meat. In *Farmers as hunters: The implications of sedentism*, ed. S. Kent, 1-17. Cambridge and New York.
- Kuchikura, Y. 1988. Efficiency and focus of blowpipe hunting among Semaq Beri hunter-gatherers of peninsular Malaysia. *Human Ecology* 16: 271-305.
- Lee, R. B. 1968. What hunters do for a living, or, how to make out on scarce resources. In *Man the hunter*, ed. R. B. Lee and I. DeVore, 30-48. Chicago.
- Marshall, L. 1965. The !Kung Bushmen of the Kalahari Desert. In *Peoples of Africa*, ed. J. L. Gibbs, Jr., 241-78. Prospect Heights, Ill.
- Polo, M. 1927. *The Book of Ser Marco Polo the Venetian, concerning the kingdoms and marvels of the East*, ed. G. Parks. New York.
- Roaf, M. 1990. *Cultural atlas of Mesopotamia and the ancient Near East*. Oxford.
- Schalk, R. 1977. The structure of an anadromous fish resource. In *For theory building in archaeology*, ed. L. R. Binford, 207-50. New York.
- Shipman, P. 1986. Scavenging or hunting in early hominids: Theoretical framework and tests. *American Anthropologist* 88: 27-43.
- Stephens, D. W. and J. R. Krebs. 1986. *Foraging theory*. Princeton, N.J.
- Trigger, B. 1965. *History and settlement in lower Nubia*. Yale University Publications in Anthropology, No. 69. New Haven, Conn.
- Welch, P. 1991. *Moundville's economy*. Tuscaloosa, Ala.

II.G.11 🦢 Geese

The common domestic geese are derived from two wild species, the greylag, *Anser anser*, and the swan goose, *Anser cygnoides*. The wild greylag is found seasonally throughout most of Eurasia and North Africa, although it is not known to breed south of 45° latitude. The swan goose is confined to East Asia, although the two can freely hybridize at the meeting point of their ranges. It has been argued that the Indian bar-headed goose, *Anser indicus*, also played a part in the evolution of domestic geese, because it is interfertile with the aforementioned species (Crawford 1984).

Geese are easily domesticated, and this process probably occurred numerous times. Moreover, the continuing presence of the greylag suggests that there was constant introgression from the wild form, accounting for Charles Darwin's observation that "the amount of variation that it has undergone, as compared with that of most domesticated animals, is singularly small" (cited in Crawford 1984: 345).

Other species of geese have been domesticated on an experimental basis in ancient or modern times. These are:

Canada goose	<i>Branta canadensis</i>
White-fronted goose	<i>Anser albifrons</i>
Egyptian goose	<i>Alopochen aegyptiaca</i>
Spur-winged goose	<i>Plectropterus gambensis</i>

However, none of these birds have attained commercial importance.

Despite their distribution, domestic geese are poorly documented across their range, especially the Chinese swan goose, which is virtually absent from all discussions of Chinese animal husbandry. Geese bones, due to their greater size and fragility, are often more comminuted in archaeological deposits than those of chickens and small game birds. Nonetheless, the lack of reports of goose bones from archaeological sites may reflect more the prejudices of archaeology than a real absence.

Geese in Domestication

Not only are geese easily domesticated, but when rendered flightless by their weight, they can be herded long distances. It has been hypothesized that the process of domestication probably began with the capture of wild specimens that were subsequently force-fed to make them too heavy to fly. The fattening of geese for their liver (foie gras) seems to have been known at least from Roman times, when birds were given a mixture of flour, milk, and honey (Zeuner 1963).

Goose eggs, in addition to goose flesh, were eaten in ancient Macedonia and Rome. Like chickens, the birds were also bred for fighting,

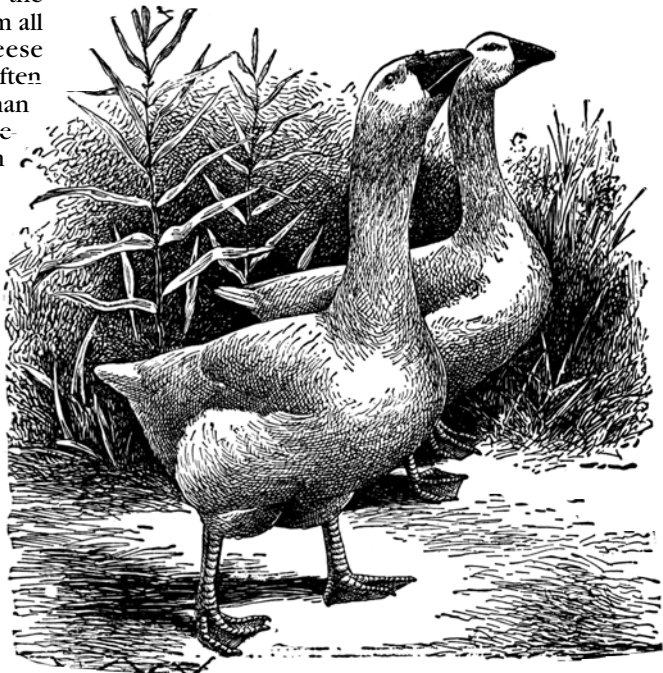
especially in Russia (Dmitriev and Ernst 1989). Geese are known in fable for raising the alarm that saved Rome from Gallic invasion in 390 B.C., and they were recorded ethnographically in Europe and South America as guard animals (Zeuner 1963; Crawford 1984). Geese are still reared for their quills, which were once of major importance for writing implements and remain sought after for some musical instruments. Geese are also kept for their down, which is used to stuff pillows, bedclothes, and clothing because of its insulating qualities.

Historical Distribution

Egypt and Southwest Asia

R. D. Crawford (1984) hypothesized that the greylag goose was probably first domesticated in southeastern Europe around 3000 B.C. However, in the absence of any tangible evidence from this region, archaeological findings now indicate Egypt as the first known center of goose domestication. The earliest evidence comes from the Old Kingdom (2723–2300 B.C.), in the form of representations of geese confined in poultry yards, kept in cages, and being herded. The pictures are sufficiently imprecise that it is uncertain whether the greylag goose or the white-fronted goose, *A. albifrons*, is being depicted. Both were trapped and eaten and, on occasion, hand-fed to increase their plumpness for the table (Houlihan and Goodman 1986: 54 ff.).

Arguably domesticated, greylag geese begin to occur osteologically in temple or funerary offerings only during the Middle Kingdom. For instance, they



Geese

have been documented from the queen's burial chambers in the Twelfth Dynasty pyramid of Amenemhat III (1850–1800 B.C.) (Darby, Ghalioungui, and Grivetti 1977; Houlihan and Goodman 1986; Katzmann 1990). Greylag geese were certainly fully domesticated by the Eighteenth Dynasty (1450–1341 B.C.), when tomb paintings showed the distinctive heterogeneous colorings of domesticated *A. anser* flocks (Boessneck 1991). If, however, the white-fronted goose (*A. albifrons*) was domesticated at that time, this practice did not persist, whereas the rearing of the greylag goose spread westward along the North African littoral and into the Near East and Europe. Despite its belligerent nature, the Egyptian goose, *A. aegyptiaca*, was also apparently domesticated for the table in the Old Kingdom, although no trace of this bird remains today (Houlihan and Goodman 1986: 64).

J. Boessneck (1991) has pointed out that the geese remains from Egypt suggest very large animals, exceeding even modern commercial breeds. These "huge domestic geese" are known from the late dynastic periods at Tell el-Maskuta (c. 600 B.C. to A.D. 200), which has yielded close to 1,000 identifiable elements. As with chickens, it may be that domestic geese were originally high-status or cult animals in Egypt and were commonly consumed only in late Dynastic or Ptolemaic times.

Geese are also represented in Mesopotamian art from the early Dynastic period onward, and it appears that in neo-Babylonian times they were farmed out to breeders on a profit-sharing basis (Zeuner 1963: 469).

Europe

The ultimate antiquity of the domestic goose in Europe cannot be effectively determined from the sparse archaeozoological literature. Domestic geese are mentioned in textual sources from the first millennium B.C. onward, as for example in *The Odyssey*, or as depicted on Macedonian coinage (Zeuner 1963: 466–7). These dates are relatively late, however, suggesting that domestic geese may not have been as ancient in Europe as many earlier scholars assumed (e.g., Zeuner 1963; Crawford 1984). Against this interpretation, various authors have noted the antiquity of the term for goose in Indo-European languages (e.g., Hahn 1926). But recent scholarship has shown that Indo-European reconstructions of terms for goose may refer just as easily to wild as to domestic birds (Gangrelize and Ivanov 1995).

Archaeozoological evidence indicates that the domestic goose did not become widespread in Europe until Roman times. Julius Caesar noted that the Britons did not eat goose because they considered it a sacred bird (Zeuner 1963). After the decline of Rome, the popularity of geese continued to increase and their spread in medieval northern Europe seems to have owed a great deal to the Normans. Geese were of minor importance in the Anglo-Saxon and Viking economies of the British Isles but

became almost central after the coming of the Normans in the eleventh to thirteenth centuries (Allison 1985; MacDonald, MacDonald, and Ryan 1993).

The economic importance of geese relative to chickens seems to have depended greatly upon local environment, with high percentages of goose bones showing up in excavations of the medieval fenland cities of East Anglia (e.g., Lincoln and Kings Lynn) (Astill and Grant 1988: 163). Likewise, geese seemed to be the most important small livestock component of the Anglo-Norman economy in parts of Ireland (e.g., Dublin) (MacDonald et al. 1993). Geese had three advantages over chickens: (1) they could be driven rather than carried to market; (2) they could thrive on food of poorer quality than that required by other domestic fowl; and (3) they would have been more significant calorically under traditional management than chickens. As a consequence of these advantages, evidence suggests that geese made a substantially greater dietary contribution in medieval towns than they did in rural areas (Astill and Grant 1988). However, this once important European food species has been displaced in recent years by the turkey, which is more amenable to modern factory farming techniques (Crawford 1984).

Africa

Sture Lagercrantz (1950: 82–7) reviewed references to geese in Africa and concluded that almost all reports of greylag geese are connected with direct European or Arab contact. It is generally assumed that the domestic goose did not cross the Sahara. It is true that the Songhay people in Mali have geese (Rouch 1954: 21), but this may be the result of relatively recent contacts with Morocco. Al-Umari mentioned that under Mansa Musa, the ruler of Mali in the fourteenth century, the peoples of the "Sudan" kept geese (Levtzion and Hopkins 1981: 267). It is possible, however, these were not *A. anser* but rather another species of goose, domesticated or tamed. The spur-winged goose, *P. gambensis*, has been recorded at San, Bamako, and Segou in Mali, where it is kept as a backyard species, as well as in northeast Nigeria (RIM 1992: 2).

Central and East Asia

The distribution and importance of the domestic form of the Chinese goose appears to be virtually undocumented. No information, either cultural or archaeozoological, is available for the swan goose, though all writers assert that it is embedded in Chinese culture. J. Delacour (1954) stated that goose rearing probably began in China more than 3,000 years ago, and William Watson (1969: 394) referred to pottery models of geese, probably representing domestic individuals, found at Hupei, a Lung Shan site in China.

The most complete documentation of geese in central Asia is in the work of N. G. Dmitriev and L. K. Ernst (1989), describing the livestock of the former Soviet Union. Originally, Chinese geese were wide-

spread in the eastern part of the nation, whereas greylags were in the west with a number of stabilized crosses developing in the center of the country. However, breeding stations have imported numerous European breeds and crossed them with indigenous stock, leading to substantial introgression in local races. Breeds were originally developed specifically for fighting and to resist extremely low temperatures.

Future Research

Information concerning the history of the domestic goose is very sparse. Obviously, an important avenue for the creation of new data on this subject rests with archaeozoology. The greatest difficulty for archaeologists with small and often comminuted assemblages is to sort out wild from domestic geese by means other than inference alone. Fortunately, various criteria have been put forward that seem to reliably differentiate wild and domestic geese in larger bird bone assemblages. These criteria include:

1. A reduction in the length and robusticity in the wing of domestic *A. anser*, coupled with an increase in lower limb robusticity relative to wild taxa (Reichstein and Pieper 1986).
2. A reduction in the wing musculature of flightless, overfed, or clipped *A. anser* with osteological effects, such as the profound reduction of feather attachments on the ulna (*papillae remigales caudales*) (MacDonald et al. 1993).
3. The presence of a flattened facet on the anterior surface of the femoral head, a pathology resulting from an overdependence on the legs for locomotion (MacDonald et al. 1993).

Apart from archaeological inquiry, ethnographic, linguistic, and textual research are also necessary, especially in East Asia. The place of the goose in Chinese culture should be reviewed through examination of iconographic and written sources as well as collation of synchronic descriptions of goose management. Virtually no work has yet been undertaken on the historical genetics of geese, making the task of tracing their diffusion across the Old World extremely problematic. These lacunae probably reflect the lack of current economic significance accorded geese rather than their historical role in human nutrition.

Kevin C. MacDonald
Roger Blench

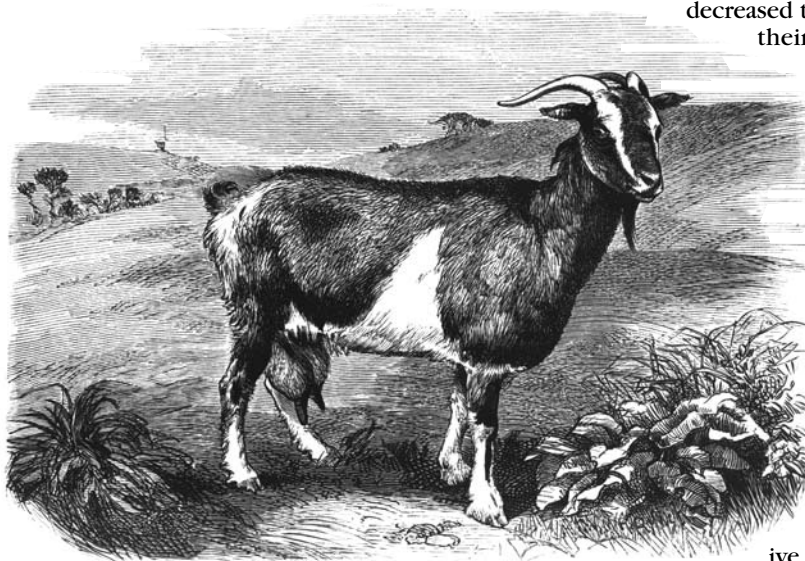
Bibliography

- Allison, E. P. 1985. Archaeozoological study of bird bones from seven sites in York. Unpublished doctoral thesis, University of York.
- Astill, G., and A. Grant. 1988. *The countryside of medieval England*. Oxford.

- Boessneck, J. 1991. Riesige Hausgänse aus der Spätzeit des alten Ägypten. *Archiv für Geflügelkunde* 55: 105–10.
- Crawford, R. D. 1984. Geese. In *Evolution of domesticated animals*, ed. I. L. Mason, 345–9. London.
- Darby, William J., Paul Ghalioungui, and Louis Grivetti. 1977. *The gift of Osiris*. 2 vols. London.
- Darwin, Charles. 1868. *The variation of animals and plants under domestication*. London.
- Delacour, J. 1954. *The waterfowl of the world*, Vol. 1. London.
- Dmitriev, N. G., and L. K. Ernst. 1989. *Animal genetic resources of the USSR*. FAO Animal Production and Health Paper, 65. Rome.
- Gamgrelize, T. V., and V. V. Ivanov. 1995. *Indo European and the Indo Europeans: A reconstruction and historical analysis of a proto-language and a proto-culture*. Part 1, *The Text*. Trans. Johanna Nichols; ed. Werner Winter. Berlin.
- Hahn, E. 1926. Gans. In *Reallexikon der Vorgeschichte*, Vol. 4, ed. M. Ebert, 167. Berlin.
- Houlihan, P. F., and S. M. Goodman. 1986. *The birds of ancient Egypt*. Warminster, England.
- Hugo, Susanne. 1995. Geese: The underestimated species. *World Animal Review* 83: 64–7.
- Ives, P. P. 1951. *Domestic geese and ducks*. New York.
- Katzmann, L. 1990. Tierknochenfunde aus Elephantine in Oberägypten (Grabungsjahre 1976 bis 1986/87) Vögel, Reptilien, Fische und Mollusken. Doctoral dissertation, University of Munich.
- Lagercrantz, Sture. 1950. *Contributions to the ethnography of Africa*, Studia Ethnographica Upsaliensia, I. Westport, Conn.
- Levtzion, N., and J. F. P. Hopkins, eds. 1981. *Corpus of early Arabic sources for West African history*. Cambridge and New York.
- MacDonald, R. H., K. C. MacDonald, and K. Ryan. 1993. Domestic geese from medieval Dublin. *Archaeofauna* 2: 205–18.
- Reichstein, H., and H. Pieper. 1986. *Untersuchungen an Skelettresten von Vögeln aus Haitbabu (Ausgrabung 1966–1969)*. Neumünster, Germany.
- Resource Inventory and Management (RIM). 1992. *National Livestock Resource Survey*. 6 vols. Abuja: Final report to Federal Department of Livestock and Pest Control Services, Federal Government of Nigeria. Abuja, Nigeria.
- Rouch, Jean. 1954. *Les songhay*. Paris.
- Schneider, K. H. 1988. *Gänse*. Berlin.
- Watson, William. 1969. Early animal domestication in China. In *The domestication and exploitation of plants and animals*, ed. P. J. Ucko, and G. W. Dimbleby, 393–5. London.
- Zeuner F. E. 1963. *A history of domesticated animals*. London.

II.G.12 🐐 Goats

The domesticated goat (*Capra hircus*) is an animal that, although of extraordinary usefulness to humans, experiences sharply different levels of acceptance around the world. Its ruminant ability to digest cellulose is the key to its success as a form of livestock, but its browsing efficiency can often be harmful to marginal environments.



Goat

Origin and Domestication

Goats were domesticated in the Near East from *Capra aegagrus*, known variously as the Persian wild goat, bezoar goat, or *padang*. The males of this ungulate species of the rugged terrain of western Asia have long, scimitar-shaped horns; the females' horns are similarly shaped but shorter. The bezoar goat has been a prey of hunters, in part, at least, because its stomach concretions (also called "bezoars" by physicians) have a widespread but medically unfounded reputation as an antidote for poison. The foothills of the Zagros Mountains is the most plausible area for the origin of goat domestication. Early Neolithic sites contain evidence of goat keeping from as long as 9,000 years ago. Such dating would seem to make the goat a candidate for the world's oldest domesticated herd animal. Brian Hesse (1982) analyzed the abundant goat bones at the site of Ganj Dareh on the cold Iranian plateau and determined that the smaller size of the bones corresponded to domestic goats. Other sites, dated several hundred years later, have yielded further evidence of early goat keeping in the eastern half of the Fertile Crescent. At Jarmo, goats were the most numerous domesticate. At Tepe Ali Kosh, domestic goats preceded domestic sheep. From the east, the domesticated caprine spread westward into the Mediterranean. For example, at Natufian sites in the Levant, the domesticated goat appeared only later as a significant animal in the life of the people. There, and in Egypt, goats have been present for about 7,000 years. In the Nile Valley, goats were integrated into a sedentary agricultural system, but they also could be sustained in the non-irrigated desert beyond.

As with sheep, the domestication process

decreased the average size of goats from that of their ancestors. Although the domestic goat bore a horned resemblance to the bezoar, smaller horns gradually evolved, and a number of polled breeds also emerged. The lop ears and twisted horns of some goat breeds reveal human selection of mutations. As with other domesticated animals, the coat colors and patterns of domestic goats are more diverse than those of their wild ancestors. Unlike sheep, goats have maintained the agility, intelligence, and curiosity of the wild animal. These features enable domestic goats to bond with humans much more strongly than do sheep and, at the same time, to be more willful and combative. As was the case with other animals,

domestication increased libido; the adjective "goatish" means "lustful." Although goats and sheep belong to the same family (Bovidae), and even the same tribe (Caprini), these two different genera do not hybridize.

Domestication may not have been for the economic reasons that seem so seductively straightforward today. The need for food was not necessarily a plausible mother of invention, for domestication is a process that occurs over many generations. A quite different motive seems more worthy of consideration. As with sheep and other domesticates, taming and reproduction in captivity may well have begun for the purpose of maintaining a steady supply of sacrificial animals. This theory is suggested by the important role that goats played in religious ritual, often as symbols of fecundity and virility. In Sumer, the deity known as "Enki" had the tail of a fish and the head of a goat. The Israelites sacrificed goats to Yahweh. The well-known figurative concept of the "scapegoat" came from the practice of priests' placing their hands on the heads of goats and recounting the sins of the people.

The importance of sacrifice probably motivated early cult masters to keep a ready supply of animals for this purpose, and controlling reproduction ensured a flow of live animals that the vagaries of hunting could not. Several other religions, such as Zoroastrianism and Hinduism, have sacrificed goats. In the latter case, goats are sacrificed to "Kali," the Hindu goddess of time, who is much revered in the Bengal region of India (Samanta 1994). A healthy male goat, black in color, is prepared for sacrifice through purification rites. Then the animal is decapitated in conjunction with the invocation of Kali, "who thirsts for blood." The blood is ritually drunk; the rest of the goat is then immolated, and sometimes the cooked flesh is eaten.

In the art of ancient Greece, satyrs were the attendants of Dionysus and Pan and had goat-like attrib-

utes. They had the legs and hindquarters of a goat, small horns, and goat-like ears. The gods themselves were goat-like: Pan had a man's body and head, but the hooves and horns of a goat; Dionysus, who assumed the form of a goat, was, like Zeus, raised on goat's milk. The cults of Zeus, Apollo, Hera, Hermes, and Artemis all manifested caprine elements. Aphrodite was sometimes depicted as riding a goat. In ancient Rome, goats were sacrificed during the feast of Lupercalia. Goat imagery of the classical period was the source of goat-like demons of the Middle Ages. Witchcraft has used the goat as a symbol of evil. The Biblical metaphor of the sheep (those chosen by God) and the goats (the damned) has doubtless influenced Western thinking about the relative worth of these two animals.

Goat Distribution

The world's pattern of goat keeping is very uneven. The animals play a major role in subsistence economies, where they provide useful products in return for a minimal investment of capital and labor. Thus, nonindustrial countries contain more than 90 percent of the half-billion goats in the world (FAO 1995). Goats adapt well to both dry and wet conditions and to tropical and midlatitude climates, and culture also helps to explain the importance the goat is accorded in the developing world. Islamic societies of the Middle East and North Africa have ennobled the goat as a source of food, and the Arab conquest of North Africa greatly expanded pastoralism as a way of life and the goat's role in it. By contrast, the Romans in North Africa emphasized farming to a much greater extent. Yet the Islamic factor must also be weighed by the ability of the goat to thrive in a subhumid environment. Thus, for example, the rural folk of both Greece and Turkey have paid much attention to goat keeping. The summer drought of the Mediterranean climate has always made it difficult to supply the large amounts of better-quality fodder needed to sustain herds of cattle. Keeping goats as an alternative avoids this constraint.

Though only partly semiarid and Islamic, Nigeria has more goats – about 250 million – than any other country in Africa. No controlled breeding is practiced, and the typical herd comprised several different kinds, including dwarf goats, which stand less than 50 centimeters high (Epstein 1971). Goats constitute about half of the domestic grazing livestock in Nigeria and provide about one-quarter of the meat consumed in that country. Goat meat supplies protein, mostly to rural people, who also sell live animals in markets as a source of income. The animals generally fend for themselves, subsisting largely on browse discovered within a radius of their dwelling compound. Sometimes they are fed vegetable waste from the kitchen, and their owners typically provide water. Night shelter within the compound protects them from preda-

tors and thieves. During the planting season, goats are often tethered to prevent crop destruction. Most of the animals have parasites, and their mortality rate from disease is high.

Goats are also part of the rural tradition in Mexico. Most of the 9 million goats there browse the sparsely vegetated slopes throughout the country. Poor Mexican farmers have long counted on their *cabras* as sources of animal protein and income. The *criollo* breed, which evolved from goats brought from Spain in the sixteenth century, is well adapted to the hot, dry landscape. However, it is low-yielding in edible products and is a carrier of brucellosis, a disease of domestic animals often transmitted to humans. Goat keeping in Mexico today has the reputation of a backward livestock activity, and environmentalists blame free-ranging animals for much of the erosion of the hill lands.

Uses of Goats

Goats provide four major products for human use: meat, milk, fiber, and skins. In nonindustrialized societies, owners typically make use of all four. In Europe, North America, and some other places, goats are normally kept only for their milk.

Meat

Goat meat is a vital addition to the diets of millions of people, although such consumption is not accurately recorded by official market statistics because a goat is often raised and butchered by the same family that consumes it. The small carcass size of a goat allows the animal to be consumed by a household in two or three days, and meat spoilage is consequently less than if a family butchered a cow.

Goat flesh is especially important in rural Africa, the Middle East, and South Asia; in India, members of a number of Hindu castes eat it with equanimity. In several Asian countries, goat and sheep meats are not differentiated, and, in fact, some languages do not distinguish between the two. By contrast, northern Europeans and North Americans generally avoid goat meat, with readily available meat alternatives as well as negative feelings about goats as human food constituting the explanation. In the United States, the flesh has a reputation for being smelly and tough, an opinion based on the cultural prejudices of those who have never eaten it. One major exception, however, is the attitude among people of Latin American origin, who have created a specialized ethnic demand for goat flesh. Superannuated dairy goats are slaughtered for this market, and in southern Texas meat goats are raised for sale (Glimp 1995).

Compared to other forms of livestock, goats have a structural disadvantage in that their higher metabolic rate makes them less efficient meat producers. A 350-kilogram (kg) cow requires the same amount of energy as six goats – each weighing only 25 kg. Moreover,

slaughtered goats have a lower percentage of useful parts and, correspondingly, more offal. That same 350-kg cow yields 189 kg of product, whereas the six goats together yield only 81 kg (McDowell and Woodward 1982).

Another problem is that goat meat tends to be dry because the animal has its fat deposited around the viscera rather than in marbled fashion, as do cattle and sheep. Age and sex are also factors. A young doe is palatable, but the tough meat from an old billy goat is not. Adult males have a scent gland that can make the strong flavor of their meat unacceptable to consumers. Kids, however, either as sucklings or when weaned, produce a meat of tender texture and delicate flavor. In France, kid is the only form of goat meat that finds a ready market.

Some goat dealers in the western United States have tried to distribute goat meat commercially under the euphemism of "chevon." Because goat meat is low in fat, producers have sought to capitalize on the national concern about fat in the diet. Sensory experiments have demonstrated that it is possible to use goat meat, alone or in combination with beef, to produce lowfat meat products for consumers who would not ordinarily consume goat meat (James and Berry 1997). More attention to breeding quality meat animals may have a long-term effect in a gradual American acceptance of goats as food. The Boer goat, originally from South Africa, is considered to be the best meat breed.

Milk

Milk is an important product of the goat, and in some areas dairy products constitute the only reason for keeping the animals. Yet from a commercial point of view, goats cannot compete with cows. Milk yields per lactation, duration of lactation, percentage of protein, persistency of milk yield, and the suitability of milking machines are all measures in which cows perform better than goats. Moreover, the need to maintain many more goats than cows for milk production entails more work in milking, more health problems, and more record keeping. However, in the overall perspective of single-family subsistence, the goat's small size fits in better with a family's needs than that of the much larger cow, and goats have a strong home instinct and require less herding than cattle. Compared to sheep, goats have a higher level of milk production, perhaps because the incidence of twinning is high in all breeds.

In Africa and western Asia, milk production is an important aspect of keeping goats. At a country level, India is the world leader in producing goat milk; other producers include Iran, Pakistan, Somalia, and Sudan. The Nubian goat, red or black in color, is an important milking breed of the Old World. It is recognized by its Roman nose, short hair, long legs, and lack of horns.

In zones of Western culture, goat's milk is now less important than in other parts of the world. In

Latin America, Mexico and Brazil are the two main producers of goat milk. *Dulce de leche*, a caramelized concoction of sugar and goat's milk, is a much-appreciated sweet in countries settled by Spaniards. In the Mediterranean region, goat's milk is locally available, especially in small villages. In Greece, about one-fourth of the milk supply is still provided by goats.

Moving north, both France and Switzerland produce substantial quantities of goat milk. The Alps have an old caprine tradition, which has given rise to seven dairy-goat breeds, the best known of which are Toggenburg, Saanen, and French-Alpine. Norway is also famous for its high-yielding dairy goats. In Western Europe outside the Mediterranean, however, goat milk is a specialty product. Fresh goat's milk is important for those allergic to cow's milk. In the United States, for example, it has been estimated that about 1 person in 1,000 has a medical need for goat milk (Haenlein 1996). Differences in the protein makeup of goat's and cow's milk make that of the former a nutritious substitute.

The American dairy-goat population of 1.5 million females is divided among six breeds. The Nubian is numerically the leading one but is surpassed in actual milk production by the three long-established Swiss breeds. California is the leading producer of goat's milk in the United States. Total production in the country is estimated at 600,000 tons, and about 300 farms and businesses sell goat's milk or goat-milk products. However, production, processing, and marketing of goat's milk is much less regulated by the states than is cow's milk. "Natural-food" stores often sell goat-milk products, including butter, ice cream, yoghurt, cosmetics, and soaps. In addition, goat milk is used to feed calves for veal production and to nurse zoo animals and pets.

Much of the world's goat milk is converted into yoghurt, butter, and cheese. Middle Eastern countries produce goat cheeses largely unknown outside their borders. Making cheeses has become the main use of goat's milk in Europe, where the relatively high price has encouraged their development as specialties for cheese lovers who enjoy their tangy flavors. When overripe, however, goat cheese can be brittle in texture, acid in smell, and ammoniacal in flavor. France still produces about 20 different kinds of goat cheese. Distinctively shaped and packaged, these cheeses are often sprinkled with a condiment like paprika or black pepper. Other European countries, especially Italy, make some goat cheese, but none of them are recognized for their quality as much as those of France. Until the late 1970s, no goat cheese was made in the United States, but by the 1990s, about 50 percent of American-produced goat milk was used to make cheese. Such cheeses - half imported and half domestic - developed a popular following and a gourmet image conveyed by use of the word "chevre" on the label.

Fleece and Skins

Goats yield other valuable products with their hair and skins. The classic Bedouin tent of southwestern Asia and northern Africa is made of goat hair. Goat fibers have also been used in pastoral societies to make carpets, brushes, and felt. High-quality goat fleece, traded internationally, derives from two specific breeds. One of these is the Angora – almost always white – which came originally from Turkey and has been known since antiquity for its lustrous fleece. Countries of western Asia export this product but do not have a monopoly. Before the Turkish Sultan decreed in 1881 that live Angora goats could not be exported, the breed had already been transported to other parts of the world. They thrive in South Africa, where they were introduced in 1838, but only much later did mohair become the main objective in raising them. At present, some 1.3 million Angora goats around Port Elizabeth in Cape Province yield 6 million kg of mohair annually.

In North America, Angora goats date from 1849 and have become highly concentrated on the Edwards Plateau of southern central Texas. More than 2 million animals are herded there on the open range, primarily to produce mohair for textile, upholstery, and wig-making industries. Mohair is more notable for its strength and luster than for its fineness of fiber. Angora goats in the United States, larger than their Turkish ancestors, are more numerous than those of any other breed.

The cashmere goat's undercoat yields a fine fiber, known in the luxury-fabric trade as cashmere. Cashmere goats live at high elevations in northern India, Tibet, western China, Kirghistan, and parts of Iran. Selection for a cold environment was a factor in evolving the quality of fleece associated with this breed.

Goatskins are a by-product of slaughter for meat. DNA analysis has found that most of the Dead Sea scrolls were written on goatskins preserved in the dry atmosphere of the hills of Palestine. In traditional cultures, goatskins have been used as containers for liquids. Water, wine, and distilled spirits have all been transported and stored in these skins, but the availability of cheap plastic has greatly diminished such usage.

Other Uses

Goat manure is another, usually subsidiary, benefit of raising these animals. At least in one place, the Rif mountain region of northern Morocco, manure has become the *raison d'être* of goat raising. Large numbers of goats are needed to provide the fertilizer for cannabis production. Marijuana and hashish from this plant enter into the international drug trade, and large quantities of goat manure – without which the soil would soon become exhausted – are needed to maintain production of this profitable crop year after year.

The goat has another use of local importance as a

biological control agent: Goats browse undesirable plants that are otherwise expensive and difficult to control. In the western United States, they keep in check aggressive, unwanted plants such as live oak brush, gambel oak, and leafy spurge.

Production Systems

Goats lend themselves well to a pastoral way of life. In the nomadic and seminomadic systems of Africa and Asia, they form a component of the mobile livestock capital, which is moved in constant search of fresh pasture. Goats are not the most appreciated animals in a pastoralist's inventory, but they serve as a better hedge against environmental uncertainty than do cattle or sheep. The Tuareg people of the Sahara keep almost 10 times as many goats as sheep, a rational behavior in their water-scarce environment. In East Africa, goats manifest considerable resistance to the disease trypanosomiasis, which explains part of their importance there.

Goats also fit into a mixed herding and farming economy to provide meat, milk, and other products. Small farms lend themselves to goat keeping. In the tropics, goats can breed year-round; in midlatitudes, breeding occurs seasonally. Goats may be tethered or allowed to range freely on communal grazing areas. They follow a nutrition strategy of choosing grasses of high protein content and digestibility to graze upon, but they will shift to browse if its nutritive value is greater, and unlike cattle, goats can maintain themselves on browse alone.

The Environmental Effects of Goats

Free-ranging goats have a reputation for overgrazing, which happens when their numbers surpass the capacity of the land to support them. Gully erosion results when they destroy the native vegetation and compact the soil with their sharp hooves. Several characteristics of goats favor abuse of the plant cover. A mobile upper lip permits them to graze as close to the ground as sheep at the same time that their browsing capacities exceed those of their fleecy cousins. Thorny plants with small leaves are no hindrance to goats. If browse is in short supply, they will defoliate entire bushes. Goats browse to the height they can reach when standing on their hind legs. They will also forage on leaves by climbing into low trees. With their sharp hooves, they paw away soil to obtain roots and subterranean stems. Allowed to roam, goats can occupy considerable space. On the average, they move 9.7 kilometers (km) per day, compared with 6.1 km for sheep and 5.3 km for cattle (Huston 1978).

It is true that goats have contributed to desertification, but they often get more blame than they deserve. The process begins when first cattle, and then sheep, deteriorate the range, after which the invading brush is suitable only for goats, which can

survive and reproduce in such environments. Thus, goats become the last alternative on eroded lands and are often seen as totally responsible for them.

Ultimately, however, destruction of vegetation by goats is a human management failure. An arid hillside combined with excessive numbers of goats will almost always mean harm for that environment. Human unwillingness to control goat density at a desired maximum, or to restrict the animals' mobility, has determined the land-use history of many areas. Humans have frequently introduced goats to islands and then left them to their own devices. Without predators, feral goat populations have exploded and destroyed ecological balances; the histories of many islands, from San Clemente, California, to St. Helena, involve goat stories.

The Mediterranean Plan, funded by the Food and Agriculture Organization of the United Nations (FAO) in the 1950s, was the first concerted international effort to deal with such problems. Impoverishment of the Mediterranean region was attributed to the environmental degradation brought on by overgrazing. We know that some of this process was begun in ancient times because classical authors discussed the goat problem two millennia earlier (Hughes 1994). The Mediterranean localities most heavily dependent on goat products were those with the most serious erosion. In the 1940s, Cyprus, which derived one-third of its milk, two-fifths of its cheese, one-fourth of its meat, and four-fifths of its leather from goats (Thirgood 1987), was deforested because of excessive goat browsing. Vegetation flourishes in the Mediterranean region where goats and sheep are restricted. In northern Morocco, the barren slopes that are grazed by goats contrast with ungrazed enclosures that maintain a luxuriant tree cover.

Daniel W. Gade

Bibliography

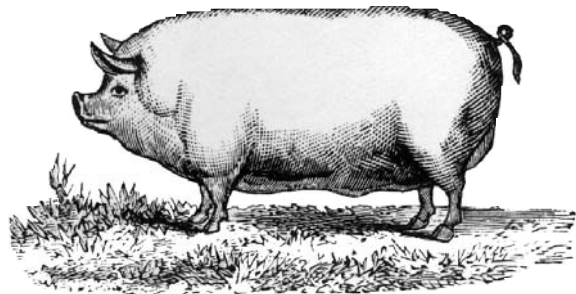
- Epstein, H. 1971. *The origin of the domestic animals of Africa*. New York.
- FAO (Food and Agriculture Organization of the United Nations). 1995. *FAO yearbook*. Rome.
- Glimp, Hudson A. 1995. Meat goat production and marketing. *Journal of Animal Science* 73: 291-5.
- Haenlein, G. F. W. 1996. Status and prospects of the dairy goat industry in the United States. *Journal of Animal Science* 74: 1173-81.
- Hesse, Brian. 1982. Slaughter patterns and domestication: The beginnings of pastoralism in western Iran. *Man* 17: 403-17.
- Hughes, J. Donald. 1994. *Pan's travail: Environmental problems of the ancient Greeks and Romans*. Baltimore, Md.
- Huston, J. E. 1978. Forage utilization and nutrient requirement of the goat. *Journal of Dairy Science* 61: 988-93.
- James, N. A., and B. W. Berry. 1997. Use of chevon in the development of low-fat meat products. *Journal of Animal Science* 75: 571-7.
- McDowell, R. E., and A. Woodward. 1982. Concepts in animal adaptation. In *Proceedings of the Third International Conference on Goat Production and Disease*, 387-91. Scottsdale, Ariz.
- Samanta, Suchitra. 1994. The "self-animal" and divine digestion: Goat sacrifice to the goddess Kali in Bengal. *Journal of Asian Studies* 53: 779-803.
- Thirgood, J. V. 1987. *Cyprus: A chronicle of its forests, land, and people*. Vancouver.

II.G.13 🐷 Hogs (Pigs)

"Pig" is a term used synonymously with "hog" and "swine" for the one domesticated suid species, *Sus scrofa domesticus*. In livestock circles, a pig becomes a hog when it passes the weight threshold of 50 kilograms. The word "swine" transcends age and sex but to many has a pejorative ring. A "gilt" is any immature version of a sow, whereas a "barrow" is a young, castrated male that can never grow up to become a boar. After a piglet has been weaned, it becomes a "shoat." Most of these terms are not used by a general public whose only encounter with this animal is in the supermarket. The meat of this often maligned beast yields some of the world's best-tasting flesh and provides good-quality protein in large amounts.

Domestication

All domesticated pigs originated from the wild boar (*Sus scrofa*) (Epstein 1984). Within that one wild species, more than 20 subspecies are known in different parts of its natural range, which has extended from the British Isles and Morocco in the West to Japan and New Guinea in the East. But where in this vast stretch of territory the first domestication occurred is still uncertain, although the earliest archaeological records (c. 7000-5000 B.C.) have been concentrated in the Middle East and eastern Mediterranean.



Pig

Indeed, the recovery of bones of domesticated pigs has been done at Jericho (Palestine), Jarmo (Iraq), Catal Huyuk (Turkey), and Argissa-Margula (Greece), as well as other sites. But bones older than any of those from these sites were uncovered in 1994 at Hallan Cemi in southeastern Turkey. There, in the foothills of the Taurus Mountains, the pig was apparently kept as early as 8000 B.C., making it the oldest known domesticated creature besides the dog. Moreover, pig keeping at this site was found to predate the cultivation of wheat or barley. Both findings contradict the long-held twin assertions that sheep and goats were the world's earliest domesticated herd animals and that crop growing preceded the raising of herd animals.

An alternative view places the beginning of swine domestication in Southeast Asia. Carl O. Sauer (1952) suggested that the pig under human control diffused from there northward to China. However, Seung Og Kim (1994) has suggested that political elites in northern China established their authority by controlling intensive pig production as early as 4300 B.C. Certainly, archaeology and cultural hubris have combined to convince many Chinese that it was their ancestors who first domesticated the pig. The Chinese ideograph for "home" consists of a character for "swine" beneath another character for "roof" (Simoons 1991).

Certain innate traits of the wild boar make it plausible that multiple domestications have occurred at different times and places in the Old World. This inquisitive, opportunistic artiodactyl may, in part, have domesticated itself by choosing to freely come into association with humans. Garbage at settlement sites provided a regular food supply, and human presence offered protection from large carnivores. Reproduction in captivity could have been initiated when captured wild piglets were tamed. Human control would have been easily accomplished, for it has been observed that the striped piglets of the wild boar behave just like the unstriped piglets of the domesticated species. The next step, unconscious selection, began the long process of evolving regional distinctions in the animal's conformation. However, the emergence of distinctive breeds, as we know them today, dates mostly from the late eighteenth and nineteenth centuries, when artificial selection was implemented on a large scale.

Religion probably lay behind the transformation of the pig from a semidomesticated status to one of greater mutual dependency with humans. In ancient Egypt, followers of Seth sacrificed pigs to that god. On the Iberian Peninsula, the granite sculptures called *verracos*, carved by Celts between the sixth century B.C. and the first century A.D., suggest that pigs might have had a religious role. In ancient Greek and Roman times, pigs were sacrificed to deities. In China, the Manchus believed that a sacrificial pig drove away bad spirits and assured good fortune. In all these

groups, the incentive of supplying live animals for cultic purposes could easily have resulted in breeding pigs toward greater dependency on humans.

Advantages of the Pig

From a contemporary utilitarian perspective, the pig is one of the glories of animal domestication. It is prolific. After a gestation period of only 4 months, a sow gives birth to an average of 10 piglets, though litter size may, on occasion, be as large as 30. Growth is rapid. In a 6-month period, piglets of 1.2 kilograms can potentially increase in weight by 5,000 percent. This growth translates into a higher return for energy invested than for other domesticated animals. Another advantage is the omnivory of pigs, which permits a wide range of food options; items that are plentiful and cheap can dominate the intake. For example, surplus crops, such as sweet potatoes in New Guinea, coconuts in Polynesia, maize in the midwestern United States, and barley in Denmark, are frequently enhanced in value because they can be fed to swine. A major disadvantage of pigs is their low ability to digest fibrous plant matter, so that, unlike ruminants, they cannot do well on cellulose alone.

The Range Pig

For most of their domesticated history, swine were kept in one of two ways: free-ranging in forests or sedentary in settlements. In neither case did they compete with humans for food, although pigs have the capacity to eat and thrive on the same nutrients. For the range pig, both plant and animal matter, on and beneath the forest floor, was sought. In Western Europe, where domesticated swine have been known since before 4000 B.C., they ate acorns, chestnuts, beechnuts, hazelnuts, and wild fruits such as berries, apples, pears, and hawthorns. Their powerful mobile snouts and sharp teeth were able to dig mushrooms, tubers, roots, worms, and grubs from the ground. Eggs, snakes, young birds, mice, rabbits, and even fawns were consumed as opportunity arose.

The use of pannage (pasturing in a forest) to feed pigs was recorded from antiquity in Europe and still has not totally disappeared. An abundant iconography suggests the role that swine played in the development of European rural society. The pig is always pictured as an animal with a long flat neck, straight back, narrow snout, small erect ears, and long legs. Nimble and resourceful, it thrived on mast (nuts from the forest floor). In the early Middle Ages, mast rights were a greater source of income from the forest than the sale of wood. But pannage required peasants to enclose their fields with wooden palisades or hedges to prevent pigs from entering and destroying their crops. As concern for forest resources grew, the pannage season was fixed by seignorial decree. The main feeding period came in the autumn, when nuts,

a highly concentrated source of nutrition, fell in large numbers. In many places, it became traditional to begin mast feeding on the feast of Saint Michael (September 29) and to conclude it on the last day of November (Laurans 1976).

Mast feeding has now disappeared from Europe except in a few places. Its best-known survival is in Spain, where the oak woodland still seasonally supports black and red Iberian swine (Parsons 1962). Although by 1990 these rustic mast-feeding breeds made up only 4 percent of the Spanish pig population, the cured pork products derived from them have been prized as especially delectable. Thus, cured hams (*jamon ibérico*) from these swine are very expensive; most famous are those from Jabugo, a meat-packing village in the Sierra Morena north of Huelva.

On his second voyage, Christopher Columbus brought the first pigs to the New World (1493). From an original stock of eight, they multiplied on the Caribbean island of Hispaniola, and many later became feral. Rounded up as needed, pigs were put on board ships bound for Mexico, Panama, Colombia, and all the islands in between. Francisco Pizarro, who had worked with swine in his youth in Extremadura, brought live pigs to the Andean highlands from Panama in 1531. The long-legged, nimble suid was well suited to move along with the expedition parties as a mobile fresh meat supply. Tropical America afforded no acorns or chestnuts, but an abundance of wild fruit, especially from palms, provided nourishment for the pigs. In semiarid zones, seed pods of leguminous trees were the common food of foraging swine.

The first pigs in what is now the United States arrived from Cuba with Hernando de Soto's expedition (1539–42) through the Southeast. Later introductions came from the British Isles, most notably to John Smith's settlement of Jamestown in 1607. A few years later, they had multiplied to several hundred head. In Virginia, and elsewhere in eastern North America, pigs fit well into the forested countryside as foragers. Abundant oak and chestnut mast in the Appalachians offered a good return in meat for almost no investment in feed or care. In late autumn, the semiferal animals were rounded up and slaughtered, and their fatty flesh was made into salt pork, which along with Indian corn was a staple of the early American diet. In the early nineteenth century, these Appalachian pigs were commercialized. The leading national and world position of Cincinnati, Ohio (often jokingly called "Porkopolis"), as a soap manufacturing center owes its origin to pigs brought there on barges for slaughter; their flesh was salted and their fat rendered into soap.

This type of hardy porker and wily beast of folk legend still survives in the Ozarks and elsewhere in the southern United States. In fact, these "razorbacks" could be descendants of those that accompanied the de Soto expedition. The explorer gave gifts of live pigs

to the Indians, and when he died in 1542 near what is now Fort Smith, Arkansas, 700 pigs were counted among his property. In addition, Ossabow Island, off the coast of Georgia, still harbors a breed of swine considered to be direct descendants of those brought by the Spaniards.

In addition to the Caribbean Islands, other uninhabited islands around the world became homes of the pig. In many cases, the animals were introduced by explorers and mariners and left to reproduce on their own. Sailors on passing ships often rounded up and slaughtered some of these feral pigs to replenish shipboard larders. Nonetheless, pigs on islands often multiplied to the point where they destroyed native fauna and flora.

In Melanesia, semidomesticated pigs still forage in the forest and are slaughtered primarily for ritual purposes (Baldwin 1983). R. A. Rappaport (1967) has explained the impressive pig feasts among the Tsembaga people of New Guinea as a societal mechanism that fulfills the need to control the size, composition, and rate of growth of the pig population. Without the periodic slaughtering of large numbers, the pigs would seriously damage gardens and crops. Rappaport's effort to understand pigs and ritual as part of a homeostatic balance became one of the landmark works of the developing subfield of cultural ecology. Whether such a pig complex makes economic sense has been debated because, in this case, the animal is not a regular source of human food. But aside from their meat, pigs must be appreciated in manifold ways: as a hedge against uncertainty (e.g., crop failure); as a negotiable store of surplus; as a source of fertilizer; and as disposers of garbage and other wastes.

The Garbage Pig

The garbage pig was essentially "presented" with its food intake, either at a fixed site or within a circumscribed area. In eastern Asia, where centuries-old deforestation and high population densities did not favor mast feeding, pig raising was long ago oriented toward consuming wastes. Important in China and Korea, at one time, was the privy pig, kept to process human excrement into flesh for human consumption. Four young pigs could derive sustenance from the waste of a family of four humans, which provided the animals with approximately 2 kilograms of human excreta and 220 grams of garbage each day (Miller 1990). In Asia, food provided by humans rather than by foraging promoted sedentary habits that, in turn, led to the evolution of several breeds with a sway-back and a dishlike face. But even the miniaturized types of Asian pigs have big appetites and large litters.

The garbage pig could also be found in ancient civilizations outside of eastern Asia. Robert L. Miller (1990) has brilliantly reconstructed the scavenging role of the pig in dynastic Egypt. But, thus far, similar

evidence is lacking for ancient Greece and Rome. In Europe, the garbage pig goes back to the Middle Ages but seems not to have been common until the fifteenth century, when the so-called Celtic pig, with white skin and pendant ears, emerged. Families fattened their pigs primarily on food scraps, and when winter neared, the animals were butchered. Their meat was cured and their fat rendered to make lard for cooking and especially for food preservation. Thus, the human diet was diversified during the cold months.

This form of pig keeping expanded as forest clearing advanced and the scale of food processing increased. Grist and oil mills generated large quantities of waste materials that could be consumed by pigs, as could the garbage from institutions like hospitals and convents. Before proper sewage disposal was implemented, many cities had swine populations to serve as ambulatory sanitation services. In medieval Paris, so many pigs were locally available for slaughter that pork was the cheapest meat. The monks of Saint Anthony – the patron saint of swineherds – were given special rights to keep pigs within the city walls. In New York City, pigs wandered the alleyways well into the nineteenth century. Naples was the last large European city to use pigs for sanitation. Neapolitan families each had a pig tethered near their dwellings to consume garbage and excrement.

Certain peasant societies still value the garbage pig as an element of domestic economy. In much of rural Latin America, pigs consume what they can find, to be later slaughtered with minimal investment in feed (Gade 1987). Lard has been an important product of pig keeping there. Frying was a cooking innovation introduced with the European conquest, and native people learned to depend on this source of animal fat. Today, however, the meat quality of these haphazardly fed animals no longer meets the health requirements of city dwellers, most of whom get their pork products through inspected channels.

Unlike sheep, whose wool may be more valuable than their flesh, or cattle that are kept for their milk or for use as draft animals, pigs have had no primary nonmeat uses. A possible minor exception has been the truffle pig, employed in France – mainly in the Périgord region – to locate the black truffles synonymous with gourmandise. A trained sow can detect from 6 meters away the smell of the unseen truffles.

Swine Distribution

World hog distribution is strongly affected by cultural and ecological factors. More than 40 percent of the world porcine inventory is in China, where density is among the highest anywhere: For every three people in China, there is one pig. Some of this swinish appeal is cultural preference, though much can be explained by lack of alternatives. Human population pressure in China does not permit the extravagance of devoting

large areas of land to grazing herbivores. Swine in China long had a niche as scavengers, scatovores, and consumers of surplus food crops.

Europe, including Russia, has about 170 million pigs, and Denmark is the only country in the world that has more pigs than people. The United States and Canada together have about 70 million of the animals, which means roughly one pig for every four people. Brazil has about 32 million head. Pigs are more important on many Pacific Islands than their total number (less than 5 million) would suggest.

The Middle East, however, is one part of the world that is largely devoid of pigs. Those that are kept generally belong to non-Muslims (such as the Coptic Christian peasants in Egypt), but some marginalized Muslims may keep pigs secretly. In humid Southeast Asia, the Islamic injunction against pigs is somewhat more nuanced, and in Indonesia, Muslims are among those who consume the products of the more than 8 million swine in that country.

In India, where the Hindu majority views all flesh foods as unacceptable elements of the diet, there are only about 10 million pigs, and in non-Islamic Africa, pigs number only around 18 million, considerably less than one might expect. But there African swine fever has periodically wiped out pig populations.

Elsewhere, the Arctic and Subarctic have historically had few pigs for quite different reasons. Very short growing seasons do not provide sufficient feed to maintain them; moreover, piglets cannot survive extremely cold winters without proper protection. Thus, in Greenland, the Norse settlements between A.D. 986 and 1400 had cattle but no pigs.

Pork

Many think that pork is the most savory of all flesh foods. The abundance and quality of fat keeps the meat from tasting dry and imparts a characteristically rich flavor. High pork consumption patterns are found in both China and Europe. In China, it is more important than all other meats combined, and use is made of all parts of the pig, including its liver, kidneys, feet, knuckles, tongue, skin, tail, and blood (Anderson 1988; Simoons 1991). Most pork cuts are also fried in lard.

In Europe, more than in China, cured pork products are favored. Germans and Slavs enjoy a range of sausages such as pork brain sausage (*Gehirnwurst*), much appreciated in Germany but unavailable commercially in the United States. In addition, Russians and Poles are especially fond of suckling pig.

Spaniards also enjoy high pork consumption and have a special fondness for cured hams and suckling pigs. *Cochonillo asado*, a strongly traditional Castilian meal, features a 1-month-old piglet fed only on its mother's milk and roasted in an earthenware dish. Spanish enthusiasm for pig meat stems in part from pork's past importance as a symbol of cultural

identity. Because Moors and Jews did not eat it, Christians saw the meat as more than simple nutrition. In sixteenth-century Spain, pork eating was an acid test faced by Spanish Moriscos and Marranos who publicly claimed conversion to Christianity. Conspicuous pork avoidance could result in an appearance before the tribunals of the Inquisition.

In North America, pork lost its preeminent position to beef in the early twentieth century. Aside from the fact that cattle ranchers were much better organized to promote their product, pork had acquired negative connotations as the main food in the monotonous diet of poor people and pioneers. At one time, pork also had unhealthy associations because of the potential for human infection by the organism *Trichinella spiralis*, carried by pigs, but federal meat inspection to certify trichina-free meat has greatly reduced the incidence of trichinosis. Still, wide knowledge of that old health risk continues to motivate cooks to make sure that pork is served only when fully cooked. Pork is considered done when it reaches an average interior temperature of 75.9° C (170° F).

Spareribs are often used in southern-style barbecue, and bacon, which comes from what the meat trade calls "pork bellies," continues as an important component of the traditional American breakfast. "Chitterlings," made from intestines, are eaten as part of the ethnic cuisine known as "soul food," but they appear in few other cases. In contrast, ham, mostly cured, has wide appeal. Virginia hams, especially those from Smithfield, have had the best reputation among connoisseurs. Fattened on peanuts and corn, Virginia porkers yield a ham that is smoked and then matured for a year.

Avoidance of Pig Meat

Despite the popularity of pork in much of the world, it can also be observed that no other domestic animal has provoked such negative reactions in so many as the pig. In Western countries, swine are commonly seen as a metaphor for filthy, greedy, smelly, lazy, stubborn, and mean. Yet these presumed attributes have not prevented pork consumption. In other cultures, however, strongly negative attitudes toward the pig have historically resulted in its rejection as human food and even, in some cases, as an animal fit to look at.

About one-fifth of the world's population refrains from eating pork as a matter of principle. Muslims, 800 million strong, form the largest block of pork rejecters because the pig is specifically named in the Koran as an object of defilement. A millennium earlier, Jews had also decided that the pig was unacceptable as a source of food. In the Bible (Leviticus), the animal is rejected because it does not meet the arcane requirements of both having a split hoof (which it has) and chewing the cud (which it does not). It is quite possible that the prophet Mohammed acquired

his conception of the pig as an unclean animal from the Jewish tradition.

The Jews, in turn, may have gotten their basic and negative idea about the pig from other neighboring peoples (Simoons 1994). Brian Hesse (1990), investigating Iron Age Palestine, found no evidence for a significant cultic role for pigs, which raises the intriguing question of whether Hebrew rules prohibited a food that no one ate in the first place. Although the Hittites of Anatolia kept pigs, they entertained negative notions about them. In dynastic Egypt, swineherds were a caste apart; pigs acquired a reputation for being unclean, and their flesh was not eaten by priests. Marvin Harris (1985) has asserted that because the pig does not fit into the hot and dry conditions of the Middle East, the Israelites banned it as an ecological misfit. Others, however, believe that the history of this fascinating taboo is more complicated than that (Simoons 1994).

Within the Christian tradition, adherents of the Ethiopian Orthodox Church do not consume pork, yet their religiously affiliated Coptic brothers in Egypt do eat it. Many Buddhists and the great majority of Hindus refuse pig meat, although in both cases this avoidance arises more from vegetarian conviction than from any explicit religious taboo toward the animal. The Mongols, those nomadic folk of central Asia whose way of life is ill suited to pig keeping, consider the pig to be a symbol of Chinese culture.

Historic reluctance to keep pigs and consume pork has also been noted in Scotland. Eric B. Ross (1983) has explained the rise of the Scottish aversion to the pig as a response to the ecological cost of keeping pigs after the decline of oak and beech forests. Because sheep raising best suited the subsequent moorland landscape, mutton became a cheaper and socially more acceptable source of animal protein.

Porcine Trends

In most industrialized countries since the 1960s, pig keeping has moved rapidly toward maximizing efficiency of production (Pond 1983). Animals are injected with growth hormones and spend their short lives within buildings in near darkness. Artificially inseminated sows (a technology that appeared in 1932) farrow in narrow steel cages. Piglets are removed early from their mother so that lactation ceases and her sexual receptivity is reactivated. In four months, another litter is produced. Piglets have their incisors removed, tails docked, ears notched, and in the case of males, their testes excised. Most pigs never reach their eighth month, although theoretically the animal can live about 15 years. Modern hog raising seeks to emulate the efficiency levels of the poultry industry. In that quest, its technological center in the United States has shifted from the Midwest to a belt of large corporate farms in eastern North Car-

olina. In 1994 alone, that latter-day "hog heaven" sent nearly 10 million animals, worth a total of about a billion dollars, to market. The next step, pork processing and trade, is normally quite lucrative, for about 75 percent of a pig's carcass can be made into salable cuts of meat. By-products, such as lard, bristles, gelatin, and cortisone, further enhance profitability.

For almost half a century, a spiced ham product high in fat has been a favorite canned food in the United States and many other countries. Spam, packaged by Geo. A. Hormel and Company of Austin, Minnesota, has a sales volume of close to 100 million cans a year. Since 1937, when it was put on the market, Hormel's Spam has served as an economical source of meat protein for millions of people. In 1991, following a major trend in the American food industry, Spam Lite was introduced; this product, however, barely meets federal requirements for fat content reduction.

Weight- and cholesterol-conscious consumers in Europe and North America have had an impact on the pork industry (Bichard and Bruce 1989). Consumer demand calls for leaner cuts, including substantial fat trimming, in supermarket meat cases. There is also a strong motivation to develop hog breeds with less fat in their muscle tissue, which normally has 5 to 7 percent fat. Pork fat is higher in unsaturated fatty acids than beef, veal, or lamb fat. On the average, one 85-gram serving of pork contains about 79 milligrams of cholesterol.

Lard, as a cooking medium, has ceded its former importance to vegetable oils. In response to this development, a shift in hog breeds toward the Landrace and the Large White has occurred. Of the 15 breeds of swine listed in the 1930 *USDA Agricultural Yearbook*, more than half have now disappeared. Less efficient breeds have also lost ground in Europe, where half of the 66 surviving breeds are now rare, and only 40 of 100 different breeds occurring in China are considered to be economically valuable (Epstein 1969).

International trade in pig meat originates overwhelmingly in Europe. In place for many decades, the single biggest flow of cured and canned pork is from Denmark to the United Kingdom. But even before that trade emerged, bacon had become an integral element of the British breakfast. The word "bacon" is derived from Francis Bacon (1561-1626), whose family crest featured a pig. Japan, although traditionally not a major consumer of pig meat, has become a significant importer of European pork products, and the well-known prosciutto (cured ham), from the Province of Parma in Italy, is now found in upscale shops around the world. Much potential movement of processed pig meat is thwarted because of the prevalence of three diseases, found mainly in underdeveloped countries: foot-and-mouth disease, hog cholera, and African swine fever. In Haiti in the early 1980s, for example, an epidemic of African swine fever killed two-thirds of the country's 1.2 million pigs. The sur-

viving swine constituted a reservoir of the disease and were slaughtered. Then, disease-free swine were imported from the United States and distributed to some of Haiti's 800,000 former pig owners.

Conclusion

The story of the pig in space and time is one of a multifaceted mutualism with humans. Its early roles, as a feeder on garbage or as a free-ranging consumer of mast in the forest, freed the pig from competing with people for the same food. Part of this mutualism was also hygienic because of the animal's capacity for disposing of wastes. Transforming the least noble of materials into succulent protein, however, also engendered enough apparent disgust to ban the animal and its flesh in certain cultures. The pig is now found around the world under vastly different circumstances. In industrialized countries, questions of efficiency and fat control have become of paramount importance, and no other domesticated animal has undergone such major changes in the way it has been kept.

Daniel W. Gade

Bibliography

- Anderson, Eugene N., Jr. 1988. *The food of China*. New Haven, Conn.
- Baldwin, James A. 1983. Pre-Cookian pigs in Australia? *Journal of Cultural Geography* 4: 17-27.
- Bichard, Maurice, and E. Ann Bruce. 1989. Pigmeat production worldwide. *Outlook on Agriculture* 18: 145-51.
- Epstein, H. 1969. *Domestic animals of China*. Farnham Royal, England.
1984. Pig. In *Evolution of domesticated animals*, ed. I. L. Mason, 145-62. New York.
- Gade, Daniel W. 1987. The Iberian pig in the Central Andes. *Journal of Cultural Geography* 7: 35-50.
- Harris, Marvin. 1985. *Good to eat: Riddles of food and culture*. New York.
- Hesse, Brian. 1990. Pig lovers and pig haters: Patterns of Palestinian pork production. *Journal of Ethnobiology* 10: 195-226.
- Kim, Seung Og. 1994. Burials, pigs, and political prestige in northern China. *Current Anthropology* 35: 119-41.
- Laurans, R. 1963. *Le porc dans la société médiévale*. Paris.
1976. Évolution de l'alimentation du porc domestique. I. L'alimentation traditionnelle. In *Le porc domestique* (Ethnozootechnie No. 16), ed. R. Laurans, 22-33. Paris.
- Miller, Robert L. 1990. Hogs and hygiene. *Journal of Egyptian Archaeology* 76: 125-40.
- Parsons, James J. 1962. The acorn-hog economy of the oak woodlands of southwestern Spain. *Geographical Review* 52: 211-35.
- Pond, W. G. 1983. Modern pork production. *Scientific American* 248: 78-87.
- Rappaport, R. A. 1967. *Pigs for the ancestors*. New Haven, Conn.
- Ross, Eric B. 1983. The riddle of the Scottish pig. *BioScience* 33: 99-106.

- Sauer, Carl O. 1952. *Agricultural origins and dispersals*. New York.
- Simoons, Frederick J. 1991. *Food in China: A cultural and historical inquiry*. Boca Raton, Fla.
1994. *Eat not this flesh: Food avoidances from prehistory to the present*. Second edition. Madison, Wis.

II.G.14 🐾 Horses

The horse represents one of the most successful outcomes of animal domestication, but for a variety of reasons it has not been widely used as a source of human food. Very little of the exacting attention given this creature over the past 5,000 years has been directed toward developing its latent meat or milk potential. Artificial selection of this nonruminant herbivore has focused on speed, strength, and configuration.

Domestication

Long before their domestication, wild horses roamed the Eurasian grasslands. They were a favorite subject of the Paleolithic cave art of western Europe, which suggests their status as a major prey species. Certain Upper Pleistocene kill sites, such as that at Solutré, France, have more horse bones than those of any other animal. Human intervention into their breeding came later than with other herd animals. Two wild horses, the tarpan (*Equus ferus gmelini*) and (Nikolai) Przhevalski's horse (*Equus ferus przewalski*), were the ancestors of the domesticated *Equus caballus*. Present knowledge places horse domestication in the grasslands of Ukraine around the fourth millennium before Christ. At Dereivka, a site of the early Kurgan culture, evidence of bit wear recovered archaeologically indicates that people rode horses (Anthony 1986). They also ate them, which is not surprising as the predecessors of these same people were avid consumers of the wild species.

Assuming present-day Ukraine to have been the center of horse domestication, the use of the animal spread westward during the next 500 years to eastern Europe, as well as eastward to the Transcaucasus and southward to Anatolia and the Mediterranean. Horse bones recovered at the site of Malyan (in Iran) that have been ingeniously analyzed reportedly show evidence of horse riding (Anthony and Brown 1989). By 2500 B.C. horses were well established in western Europe. Their main prehistoric role was as pullers of wheeled conveyances and as riding animals, and S. Bökönyi (1984) asserts that horses were used to pull carts before they were ridden. Horses were also eaten; in fact, the flesh of equids was an acceptable food in most societies that adopted them during the first 3,000 years of their domesticated state. Bronze Age sites in eastern Europe have yielded limb bones

broken for the marrow and brain cases cracked open to extract the brain.

The Kurgan people, along with other early Indo-Europeans, also sacrificed horses to honor the dead and propitiate the gods (O'Flaherty 1987; Dexter 1990). This practice was widespread from Europe to South Asia, where the Aryans of India made a strong religious place for the horse. Its role as human food was not very important, although horseflesh was consumed in connection with the *asvamedha*, a sacrifice of horses borrowed from an ancient Indo-European fertility rite (Bhawe 1939). It is especially intriguing to consider that the horse gained ascendancy as a useful and ritually significant animal with the migration of Indo-European people. The Hittites of Anatolia, for example, introduced a higher level of horse keeping than had been there before, and Mediterranean antiquity extended the high prestige of the horse, but with much less ritual killing. The ancient Greeks offered horses in sacrifice to Poseidon, the god of horses, but did not generally consume the flesh. The only Roman horse sacrifice honored Mars.

Horsemeat Avoidance

In the modern world, most people and cultures have rejected the flesh of horses (and its equine relatives, the mule and the donkey) as unfit to eat. But the reasons for avoidance are not necessarily the same everywhere. In some places, the horse is a rare, even absent, animal, so that people have had little opportunity to find out what they were missing. Aside from Iceland, few horses have been kept in the Arctic or Subarctic, primarily because of the large effort needed to store a sufficient amount of fodder to get the animal through long winters. The tropics have not had big horse populations, either. Tropical forests neither suit the horse's food requirements nor favor its physiological well-being. Moreover, in much of Africa, the presence of trypanosomiasis has excluded horses from large areas.

In most parts of the world where the horse is found, neither its meat nor its milk is used. Part of this avoidance can be attributed to religious injunction. Jews, Muslims, Hindus, and, at one time, Christians have all proscribed horsemeat from the diet as a badge of their faith. Orthodox Jews include it as one of the forbidden items in their Levitical list of animals because horses neither chew the cud nor have cloven hooves. Muslims are enjoined in the Koran to eschew this flesh, a prohibition that, perhaps, grew out of the social context of Arab desert oases, where the horse was a luxury animal owned by sheikhs. Horsemeat is not eaten by most Hindus (except for untouchable groups) because it is a flesh food in a primarily vegetarian culture.

Marginalization of horsemeat in Europe had a religious basis. As Christianity spread through Teutonic lands, clerics regarded hippophagy (eating horseflesh)

as a pagan residue that was tinged with barbarism. In prehistoric Ireland, horseflesh was eaten in ritual contexts, a practice that survived into the early Middle Ages. A chronicler in the twelfth century described such an occasion in which priestly sexual intercourse with a mare, sacrifice, and consumption of the flesh were involved (Giraldus Cambrensis 1982). Such behavior, outrageous to Christians, explains why horsemeat had such a bad reputation as the continent became more fully Christianized. Boniface, missionary to the Germans, undertook what amounted to a personal lobbying campaign to, successfully as it turned out, persuade the pope to place horseflesh into the forbidden category. Such a ban was certainly not based on health or scripture, but horsemeat, nonetheless, became the only specific food outlawed in the history of Christianity. Today, although most people in the Christian tradition still avoid the flesh of equines, its rejection no longer has much to do with a religious prohibition. Even the *Catholic Encyclopedia* makes no effort to recall this old taboo that had been part of canon law. Rejection can now be attributed mainly to fear of the unfamiliar. The status of the horse as an intelligent companion of humans has surely worked against experimentation with consuming its flesh, except in periods of severe food shortage.

Marvin Harris (1985) has argued that underlying these prohibitions was a stark ecological fact: The horse was an inefficient converter of grass to meat, and when compared to cattle, horses have a higher metabolic rate and a gestation period lasting two months longer. However, as H. B. Barclay (1989) points out, horses also have various advantages over cattle. They enjoy a longer life span and have greater stamina and endurance. In winter pastures, horses, unlike cattle, can paw beneath the snow to graze. Moreover, in central Asia, mare's milk yields are as high as those of cows.

In spite of early religious and later social reprobation, Europe did undergo a hippophagy movement that to some extent changed attitudes toward this meat and, as such, represents a notable case of how a food taboo broke down. Hippophagic experimentation in Europe was widespread around the middle of the nineteenth century, when conscious efforts were made to break with the old prejudice against selling and eating horseflesh. Denmark legalized its sale in 1841, as did the German state of Württemberg; Bavaria followed in 1842, and Prussia in 1843. Other countries (Norway, Sweden, Austria, and Switzerland) also legalized its sale. Russia, where horsemeat has had a historic culinary role, never banned the sale of horsemeat in the first place.

Horsemeat Consumption in Europe

In the European history of horsemeat consumption, France and francophone Belgium stand out. There the cultural barrier to selling and eating horsemeat was

more successfully breached than in any other part of Europe. As with so many French attitudes, it began with elite exhortation, but it also occurred in a national culture charmed by culinary diversity and invention. Its beginnings can be traced to governmental decisions that French people should eat more flesh foods. By 1850, the opinion was that the French person's ability to perform industrial work was hindered by insufficient quantities of nitrates in the diet. Ireland was pointed to as a negative case of how dependence on the potato and little or no meat had led to physiological degeneration. In 1853, dietary regulations set a minimum daily intake of 100 grams of meat for a small adult and 140 grams for a large person. In 1889, a government medical commission increased these thresholds to 120 and 200 grams, respectively. In the French Army, the meat ration was increased in 1881 from 250 to 300 grams per day. Regulations about meat also affected institutions. In 1864, hospices and hospitals were directed to provide meat at meals twice a day, preferably roasted, which was judged to be the most nutritious form of preparation.

As meat consumption in France rose, provisioning became an issue. Since horseflesh was then much cheaper, it was judged by the authorities to be especially worthy of adoption among lower-income groups. A group of distinguished French intellectuals and scientists, led by Geoffrey Saint-Hilaire, made a conscious effort to integrate horsemeat into the French diet. Their aim was to show that *chevaline* (as it was called) was nutritious and that the old prejudices had no rational basis. Even the hierarchy of the French Church did not interfere in this conscious rejection of a medieval taboo, and the first specialized horsemeat shop opened in Paris in 1866. Shortly thereafter, dire food shortages associated with the siege of Paris gave many of that city's inhabitants an opportunity to eat horsemeat for the first time.

Two prime factors led to a generalized acceptance of horsemeat in France. First, until the 1920s, horsemeat cost half that of comparable cuts of beef and therefore was within reach of the working class. Second, horsemeat got much publicity as a high-energy food and as ostensibly valuable in physiologically reducing the human temptation to overindulge in alcoholic beverages. In touting horsemeat, the bourgeoisie indicated their conviction that it had an appropriate role to play in the diet of the working class, whose members were the main abusers of alcohol in France.

Two basic kinds of horseflesh later characterized its consumption in France. Red meat, preferred by many for its putative association with high iron content, comes mainly from older horses. A colt (*poulain*) yields light meat, favored for its tenderness and mild flavor approaching that of veal. Patterns of consumption have maintained a certain profile through the decades, with the typical horsemeat consumer in France tending to be young rather than old,

working or middle class rather than rich, and urban rather than rural.

Hippophagy in France is highest in the northern part of the country, with southern Belgium an extension of that consumption pattern. Early in the twentieth century, large workhorses were numerous in this part of Europe where the Percheron and the Belgian breeds evolved. Consequently, Paris has had the most avid hippophagists in France, with a per capita consumption three times higher than that of western France and almost seven times higher than that of eastern France. Horsemeat was first promoted in Paris, and it was there that a specialized horse abattoir was set up from which horsemeat was wholesaled elsewhere. The Paris region has also had France's largest concentration of poor people, who became the most enthusiastic consumers of equine products. Paris, Lille, and other northern cities were major markets for red meat from aged animals, whereas hippophagists in Lyon and Grenoble preferred *vienne de poulain*.

In spite of considerable commercial activity, horsemeat in France has had a marginal position among meat products. Except during the war years, its consumption has never amounted to more than 5 percent of that of beef. Close to 60 percent of French people do not consume any horsemeat, and of the minority that do, the per capita consumption amounts to only about 5 kilograms per year. One result, but also a cause, of such a pattern is availability. Most French towns have not had a clientele large enough to support a *boucherie chevaline*, and until the 1980s, butchers selling beef could not also sell horsemeat. This prohibition was in the form of a law that had been promulgated to prevent misrepresentation of one meat for another.

Organoleptic perception also affects consumption. Many people, for example, have stated they dislike the flavor of horsemeat. However, disagreement on whether it has a sweetish taste, some other taste, or is simply tasteless suggests differences in the kinds of slaughtered animals or preconceived notions of the meat's characteristics. The appeal of horsemeat may be culturally conditioned. For example, an organoleptic evaluation in the United States of several kinds of lean beef, regular beef, lamb, horse, and horse mixed with beef fat ranked the last two lowest in appearance, smell, and taste (Pintauro, Yuann, and Bergan 1984).

Matters of health and price explain at least part of the drop in per capita consumption of horsemeat from the 1960s to the 1990s. For many years, horsemeat was valued as a particularly healthful food due to an iron content higher and a fat content lower than beef. *Chevaline* reputation suffered, however, in 1967, when a highly publicized salmonella outbreak in a school cafeteria was traced to horsemeat patties. When findings emerged that fresh horsemeat deteriorated more rapidly than beef, the French government banned horsemeat for institutional use. In addition,

awareness of *chevaline*'s possible health risk was paralleled by its rising cost. Horsemeat of the first half of the twentieth century was essentially a by-product of animals slaughtered after a useful life of pulling plows or wagons. For that reason, horsemeat was cheap compared to beef. But by the 1970s most horsemeat in France was imported, which influenced its retail price structure.

Thus in the 1990s, the French consumer paid normally as much, if not more, for horsemeat as for beef. This is because sources of horsemeat supply for the French and Belgian markets changed substantially after 1960, when mechanization of agriculture led to a sharp decline in the horse population. To meet the demand in the face of inadequate domestic sources of *chevaline*, live animals or chilled meat had to be imported from Spain, Portugal, and Eastern European countries. Between 1910 and 1980, imports of horses or horse products increased almost three times. European horses were supplemented with those from the Americas, and beginning in the 1970s, the United States became a major supplier for the French market, as did Australia and South America - especially Argentina, with its substantial horse population.

Horses raised specifically for meat production could, in theory, supply the needed tonnages and at the same time ensure quality. Colt meat in France is now mostly produced that way. Fed a controlled diet, horses grow very quickly, although they cannot compare with ruminants as an efficient converter of herbaceous matter. Breeding of the horse as a meat animal has another value of conserving the genetic patrimony of the famous work breeds of horses that are in danger of eventual extinction. In the long run, such activity might also lead to the emergence of a special kind of high-quality meat animal in the way that Hereford or Aberdeen Angus were developed as cattle with distinguishable phenotypes and genotypes.

Horses as Food in Other Countries

Several other European countries, notably Italy, Spain, Germany, and Switzerland, have achieved a small measure of hippophagic acceptance. Search for protein during the war and for some years after that accustomed Europeans to think of alternative possibilities. As recently as 1951, when meat was still scarce, the British consumed 53,000 horses. However, as other meats became abundant, horsemeat faded as a food for people. Moreover, animal rights activists in the United Kingdom have often shown vigilance about protecting the horse from becoming steak on the table.

In Asia, horsemeat is an important food in Mongolia, where the horse has been deeply integrated into a still traditional rural society. In addition, Mongolians consume mare's milk, usually in fermented form (kumiss). Though their religion bars it, Islamic people in Inner Asia have also been known to eat horseflesh.

In Japan, horsemeat is a favorite meat in the preparation of teriyaki, and in the 1980s Japan imported about one-fourth of the tonnage in the world trade of refrigerated or frozen horsemeat. Australia is a major supplier.

In the Western Hemisphere, no national culture has integrated hippophagy into its dietary possibilities, least of all those with a strong orientation toward carnivory. Argentines and Uruguayans, who are mostly of European origin, have rejected it, although some Indians of the pampas ate the meat of horses descended from those brought by the Spaniards. Some North American Indian groups who acquired the horse from the Europeans also ate horseflesh. In the United States, consumption of horsemeat is low; in fact, throughout North America, it is readily available (though not widely consumed) only in the Canadian province of Québec. French immigrants are reported as a major clientele of the 10 horsemeat shops in the city of Montreal in 1992.

The United States is the leading horsemeat-producing country in the world (FAO Production Yearbook 1990). In 1992, nearly 250,000 horses were slaughtered in the country. Thoroughbred race horses, including champions that once got high stud fees, usually end their days at the slaughterhouse. The \$500 to \$1,000 sale becomes a way to recover something of the high expense of raising them. Commercialization of this product has been both for shipment abroad as human food and for dog food.

In spite of the abundance of horses, equine flesh has never gained acceptance as a human food in the United States. Harris (1985) has concluded that horsemeat did not catch on because beef and pork could be produced more cheaply. More likely, preconceptions about the suitability of horsemeat have stemmed from an earlier bias brought by Europeans. Moreover, a small but vocal minority of Americans are actively opposed to its use as human food. One emotionally based argument is that it demeans such a noble animal to undergo slaughter as human food. Curiously, much less is said about the appropriateness of horsemeat as dog food, suggesting an ethic that requires humans to deal with horses in narrowly defined ways. Horse abuse in the United States receives much more publicity than mistreatment of cattle, sheep, or pigs.

The lack of horsemeat availability in American meat shops says much about the low demand, but proprietors are also afraid of losing business by making it available in the meat case. Episodes over the years of fraudulent labeling of horseflesh as beef have led to the (erroneous) assumption that the former is, by definition, an inferior meat, perhaps even tainted in some way. A slaughterhouse owner in South Carolina, who has long shipped horsemeat to the European market, has made several attempts to introduce horsemeat in his state with no lasting success. As was the case initially in Europe, a lower price for horse-

meat would seem to be the critical mechanism in its acceptance. Yet economics dictate that good-quality horseflesh cannot be sold at a price much below that of beef. An unusually strong American concern for healthful and lowfat foods could possibly give it a niche. But up to now its best-known usage has been at the dining service of the Harvard Faculty Club in Cambridge, Massachusetts. There it has been maintained as a quirky tradition since World War II, when it was difficult to obtain beef.

Daniel W. Gade

Bibliography

- Anthony, D. 1986. The "Kurgan culture," Indo-European origins, and the domestication of the horse: A reconsideration. *Current Anthropology* 27: 291-313.
- Anthony, D., and D. Brown. 1989. Looking a gift horse in the mouth: Identification of the earliest bitted equids and the microscopic analysis of wear. In *Early animal domestication and its cultural context*, ed. P. Crabtree, D. Campana, and K. Ryan, 99-116. Philadelphia, Pa.
- Barclay, H. B. 1989. Professor Harris, horses, and hippophagists. *Anthropos* 84: 213-19.
- Bhawe, S. 1939. *Die Yajus des Asvamedha; Versuch einer Rekonstruktion dieses Abschnittes des Yajurveda auf Grund der Überlieferung seiner fünf Schulen*. Bonner orientalistische Studien 25. Stuttgart.
- Bökönyi, S. 1984. Horse. In *Evolution of domesticated animals*, ed. I. L. Mason, 162-73. New York.
- Dexter, M. R. 1990. The hippomorphic goddess and her offspring. *Journal of Indo-European Studies* 18: 285-307.
- FAO production yearbook. 1990. Rome.
- Gade, Daniel W. 1976. Horsemeat as human food in France. *Ecology of Food and Nutrition* 5: 1-11.
- Giraldus Cambrensis. 1982. *The history and topography of Ireland*, trans. John J. O'Meara. Harmondsworth, England.
- Harris, Marvin. 1985. *Good to eat: Riddles of food and culture*. New York.
- O'Flaherty, W. D. 1987. Horses. In *The encyclopedia of religion*, ed. M. Eliade, 463-8. New York.
- Pintauro, S. J., K. Yuann, and J. G. Bergan. 1984. Nutritional evaluation of horsemeat. In *Protein: Nutritional quality of foods and feeds*, ed. M. Friedman, 187-96. New York.
- Robelin, J., B. Boccard, W. Martin-Rosset, et al. 1984. Caractéristiques des carcasses et qualités de la viande de cheval. In *Le cheval: Réproduction, sélection, alimentation, exploitation*, ed. R. Jarrige and W. Martin-Rosset, 601-10. Paris.
- Rossier, E. 1984. État actuel de la production et de la consommation de viande chevaline en France. In *Le cheval: Réproduction, sélection, alimentation, exploitation*, ed. R. Jarrige and W. Martin-Rosset, 491-508. Paris.
- Simmonds, P. L. 1885. *The animal food resources of different nations*. London.
- Simoons, F. J. 1994. *Eat not this flesh: Food avoidances from prehistory to the present*. Second edition. Madison, Wis.
- Vigarello, G. 1993. *Le sain et le malsain: Santé et mieux-être depuis le Moyen-Age*. Paris.

II.G.15 🦋 Insects

In the title of a delightful little book published in 1885, Vincent Holt asks, *Why Not Eat Insects?* The “why not” is hard to explain logically, nutritionally, or on the basis of the sheer abundance of these creatures. However, for Europeans and North Americans the eating of insects, or entomophagy, is considered a curiosity at best. And for many the idea is downright repulsive. Insects as food are found only in cartoons, or perhaps on the odd occasion when suitably disguised under a layer of chocolate. Yet for these same people, other invertebrate animals, such as oysters, snails, crayfish, and lobsters, are not only accepted as food but even viewed as delicacies.

In many other parts of the world, however, insects are considered good to eat and are appreciated for their taste as well as nutritional value. Some, like the giant queen ants (*Atta* sp.) of Colombia, are prized as delicacies and supposedly function as aphrodisiacs as well. Others, like the mompani worms of Africa, are frequently included in the diet and much enjoyed. Still others, like the cock chafer grubs of Ireland, although not much esteemed, have been used when more desirable foods were not available.

The purpose of this chapter is to present an overview of the role of insects in the diet in different parts of the world and in different time periods. We first review the use of insects as food within major geographical areas, in the present as well as in the historic and prehistoric past when information is available (a comprehensive list of insect species used as food is provided in Table II.G.15.1). We then summarize general patterns of insect use and provide information on the nutritional value of some commonly consumed insects.

Insects Around the World

The Americas

North America. The use of insects as food is generally frowned upon in North America, notwithstanding the repeated attempts by entomologists to make them more appealing. One of the best known of such attempts is R. L. Taylor's 1975 book, *Butterflies in My Stomach*, and the accompanying recipe guide, *Entertaining with Insects* (1976). The reasons that most North Americans do not include insects in their diets have to do with food preferences rather than nutritional value. The dietary exclusion of insects is sometimes justified by the assumption that many insects carry diseases and therefore eating them is dangerous, but there is little evidence to support such a notion (Gorham 1979).

The prohibition against eating insects among many contemporary North Americans can probably be traced to Europe. European colonists who settled in North America also had strong taboos against con-

suming insects. In fact, they originally classified the New England lobster as an insect and refused to eat it. But that was in the 1600s. Subsequently, the lobster has, of course, passed from the status of prohibited food to delicacy.

Native Americans, however, did not have the same prejudices against insects as food, and there are a number of ethnographic and historical accounts indicating their importance in the diet. The bulk of these reports are from the western United States, where the most heavily consumed insects were the larvae of the Pandora moth, grasshoppers, and Mormon crickets – insects that occur in great abundance during certain seasons and in certain places in the region.

Pandora moth larvae (*Colorado pandora lindseyi*) are large caterpillars found on a certain pine tree species, *Pinus jeffreyi*. They were reportedly collected in large numbers by Native American peoples, either by hand or by trenching around the bottom of trees when caterpillars descended to pupate in the soil (Fowler and Walter 1985). As recently as 1981, Pandora moth larvae were still being collected (Fowler and Walter 1985).

Grasshoppers were also frequently consumed by early Americans. They occurred in large swarms or “plagues” in drier areas at densities of more than 12,000 insects per square mile in California (Brues 1946). On the beaches of the Great Salt Lake in Utah, grasshoppers washed up by the lake can pile up in long windrows as much as 20 centimeters thick, already salted and sun-dried (Madsen 1989). The return rates for collection in kilocalories (kcal) per hour are greater than for any other known collected resource (Madsen and Kirkman 1988). Grasshoppers are relatively high in some B vitamins and can be stored (dry) for long periods in the form of flour (Sutton 1988).

Crickets are another valuable food resource; they were often collected in large quantities in trenches and stored for use in the winter (Sutton 1988). One of these was the Mormon cricket (*Anabrus simplex*), so named because Mormons tried to kill these crickets off in the late 1800s and in the process caused animosity between themselves and Paiutes (Madsen and Kirkman 1988). There is some indication that insects formed an increasingly substantial part of the Native American diet as the local habitat was modified by European newcomers (Sutton 1988).

Other insects used by the first Americans were shore flies (*Hydropyrus bians*), available in great abundance in some locations, and ants that were made into flour soup (Sutton 1988). Judging by historical accounts, grasshoppers and crickets were usually roasted and ground with pine nuts, grass seeds, or berries to make cakes, which could then be sun-dried and stored. These were called “desert fruitcakes” and (interestingly) were considered a delicacy by European travelers as well as by Native Americans (Madsen 1989).

Table II.G.15.1. *Numbers of species of insects used as food by stage of life cycle and geographic region*

Order/Family	Stage	NA	CA	SA	Afr	PI	Aus	Eur	ME	SEA	EA
Anoplura (sucking lice)											
Pediculidae	a	1	1	1	1	1	1	0	0	0	0
Coleoptera (beetles)											
Bruchidae	l, p, a	2	0	1	0	0	0	0	0	0	0
Buprestidae	l, a	0	0	1	2	0	0	0	0	1	1
Cerambycidae	l	6	7	3	6	5	7	1	1	1	0
Cicindelidae	a	0	2	1	1	1	2	0	0	0	1
Citheroniidae		0	0	0	1	0	0	0	0	0	0
Crysomelidae	l	0	1	0	0	0	0	0	0	0	0
Curculionidae	l	1	3	4	7	5	0	0	0	3	1
Dytiscidae	l, a	1	0	0	1	0	0	0	0	7	8
Elateridae	l	0	0	0	1	0	0	0	0	0	1
Elmidae	a	0	0	2	0	0	0	1	1	1	5
Histeridae	l	0	1	0	0	0	0	0	0	0	0
Hydrophilidae		0	0	0	0	0	0	0	1	1	4
Lucanidae	l	0	1	0	1	0	0	1	1	0	0
Passalidae	l	0	2	2	3	0	0	0	1	0	0
Pentatomidae	a	0	0	0	0	0	0	0	4	0	0
Scarabaeidae	l, a	1	3	2	12	6	2	7	0	10	0
Tenebrionidae	l	0	0	1	1	0	0	1	1	0	1
Diptera (flies)											
Calliphoridae	l	0	0	0	0	0	0	0	0	0	1
Culicidae	a	0	0	0	2	0	0	0	0	0	0
Ephydriidae	l, p	2	2	0	0	0	0	0	0	0	0
Leptidae	a, e	1	0	0	0	0	0	0	0	0	0
Muscidae	l	0	1	0	0	0	0	0	0	0	0
Oestridae	l	1	0	0	0	0	0	0	0	0	0
Sarcophagidae		0	0	0	0	0	0	0	0	0	1
Stratiomyidae		0	0	1	0	0	0	0	0	0	0
Tipulidae	a, l	1	0	0	0	0	0	0	0	0	0
Ephemeroptera (mayflies)											
Baetidae	a	0	0	0	1	0	0	0	0	0	0
Ephemeridae		0	0	0	0	0	0	0	0	1	1
Hemiptera (bugs)											
Belostomatidae	a, n	0	3	0	0	0	0	0	2	1	1
Coreidae		0	1	0	0	1	0	0	0	0	0
Corixidae	a	0	5	0	3	2	0	0	0	0	0
Lygaeidae	a	0	0	0	0	1	0	0	0	0	0
Naucoridae	a	0	0	0	0	0	0	0	0	2	0
Nepidae	a	0	0	0	1	0	0	0	0	1	0
Notonectidae	e, n, a	0	1	0	0	0	0	0	0	0	0
Pentatomidae	n, a	0	8	0	1	0	0	0	0	0	2
Homoptera (cicadas, hoppers, aphids)											
Cicadidae	a	3	3	1	2	2	1	3	1	4	1
Cicadellidae		0	0	0	0	0	0	0	0	0	2
Coccidae		0	0	0	0	0	0	0	0	0	3
Fulgoridae		0	0	0	1	0	0	0	0	0	0
Membracidae	n, a	0	3	1	0	0	0	0	0	0	0
Psyllidae		0	0	0	0	0	4	0	0	0	0
Hymenoptera (ants, bees, wasps)											
Apidae	e, l, p	0	3	0	2	4	0	0	1	2	2
Cynipidae	a	0	0	0	0	0	0	1	0	0	0
Formicidae	l, p, a	7	9	5	2	1	13	0	1	3	0
Meliponidae	e, l, p	0	6	4	8	3	3	0	0	2	0
Scoliidae	l, p	0	0	0	0	0	0	0	0	0	1
Tiphiidae		0	0	0	0	0	1	0	0	0	0
Vespidae	i, e, l, p	2	13	8	0	0	0	0	0	4	4
Xylocopidae	l, p	1	0	0	0	1	0	0	1	2	0
Isoptera (termites)											
Odontotermitidae		0	0	0	0	0	0	0	0	1	0
Rhinotermitidae	a	1	0	0	0	0	0	0	0	0	1
Termitidae		0	0	0	5	19	6	2	2	1	0
Lepidoptera (butterflies and moths)											
Bombycidae	p	0	0	0	5	0	0	0	1	1	1
Citheroniidae	l	0	0	0	1	0	0	0	0	0	0
Cossidae	l	0	2	0	0	0	5	0	0	3	0
Geometridae	l	0	1	0	0	0	0	0	0	0	0

(continued)

Table II.G.15.1. (Continued)

Order/Family	Stage	NA	CA	SA	Afr	PI	Aus	Eur	ME	SEA	EA
Lepidoptera (butterflies and moths)											
Hepialidae	l	0	1	0	0	0	2	0	0	0	2
Hesperiidae	l	0	1	1	0	0	0	0	0	0	0
Hyblacidae	l	0	0	0	0	1	0	0	0	0	0
Lacosomidae		0	0	1	0	0	0	0	0	0	0
Lasiocampidae	l, p	0	0	0	3	0	0	0	0	0	0
Limacodidae	l	0	0	0	1	0	0	0	0	0	1
Megathymidae	l	0	1	0	0	0	0	0	0	0	0
Noctuidae	l	1	4	3	1	0	1	0	0	0	0
Notodontidae	l	0	0	1	9	0	0	0	0	0	0
Pieridae	l	0	1	0	0	0	0	0	0	0	0
Psychidae	l, p	0	0	0	2	0	0	0	0	0	0
Pyralidae	l	0	1	0	0	0	0	0	1	0	0
Saturniidae	l, p	2	3	1	29	0	0	0	1	0	1
Sphingidae	l	1	0	0	1	0	0	0	0	0	0
Unknown	l	1	0	0	1	0	0	0	0	0	0
Neuroptera (alderflies, dobsonflies, lacewings)											
Corydalidae		0	0	1	0	0	0	0	0	0	1
Odonata (dragonflies)											
Aeschnidae	n, l	0	1	0	1	1	0	0	0	1	1
Libellulidae	l	0	0	0	0	1	0	0	0	0	0
Orthoptera (grasshoppers, crickets, mantids, and cockroaches)											
Acrididae	a, n	10	18	8	10	4	1	2	6	3	5
Blattidae		0	0	0	0	0	1	0	0	2	2
Gryllacrididae	a	0	0	0	0	1	0	0	0	0	0
Gryllidae	a	0	0	0	2	2	0	0	1	3	2
Gryllotalpidae	a	0	0	0	1	2	0	0	1	1	1
Mantidae	a	0	0	0	1	1	0	0	0	2	1
Phasmatidae	a	0	0	0	0	2	0	0	0	2	0
Phylliidae	a	0	0	0	0	1	0	0	0	0	0
Pyrgomorphidae	a, l, p	0	0	0	1	0	0	0	0	0	0
Tettigoniidae	a, p	1	0	1	2	1	0	3	1	1	0
Unknown	a	0	0	0	1	0	0	0	0	0	0
Plecoptera (stoneflies)											
Pteronarcidae	a	1	0	0	0	0	0	0	0	0	0
Siphonaptera (fleas)											
Unknown	a	0	0	0	0	1	0	0	0	0	0
Trichoptera (caddisflies)											
Hydropsychidae	l	0	1	1	0	0	0	0	0	0	0
Totals		48	116	56	136	70	120	22	30	67	60

Stage of life cycle: adult (a), egg (e), immature stage (i), larval (l), nymph (n), pupal (p). Geographic regions are North America (NA), Central America and the Caribbean (CA), South America (SA), Africa (Afr), Pacific Islands (PI), Australia and New Zealand (Aust), Europe (Eur), Middle East and South Asia (ME), Southeast Asia (SEA), East Asia (EA). Each named species was counted as one, unnamed species designated simply as spp. were also counted as one species.

References

North America (Aldrich 1912a, b; Bodenheimer 1951; Engelhardt 1924; Fenenga and Fisher 1978; Madsen and Kirkman 1988; Sutton 1985; Sutton 1988; Swezey 1978; Taylor 1975); Central America and Caribbean (Ancona 1933; Bequaert 1921; Bodenheimer 1951; Bristowe 1932; Conconi and Bourges 1977; Conconi et al. 1984; Fladung 1924; Holt 1885; Pierce 1915; Taylor 1975; Yturbide 1986); South America (Bequaert 1921; Bodenheimer 1951; Conconi and Bourges 1977; Dufour 1987; Fladung 1924; Ruddle 1973; Taylor 1975; Wallace 1853); Africa (Bequaert 1921; Bodenheimer 1951; Brickey and Gorham 1989; Chavanduka 1975; Conconi and Bourges 1977; DeFoliart 1990; Fladung 1924; Harris 1971; Hoffman 1947; Oliveira et al. 1976; Quin 1959; Taylor 1975; Turk 1991); Pacific Islands (Bodenheimer 1951; Bristowe 1932; Conconi and Bourges 1977; DeFoliart 1990; Holt 1885; Meyer-Rochow 1973; Taylor 1975); Australia and New Zealand (Bodenheimer 1951; Conconi and Bourges 1977; Fladung 1924; Taylor 1975); Europe (Bodenheimer 1951; Conconi and Bourges 1977; Holt 1885; Pierce 1915); South Asia and the Middle East (Bequaert 1921; Bodenheimer 1951; Conconi and Bourges 1977; Fladung 1924; Hoffman 1947; Pierce 1915); Southeast Asia (Bodenheimer 1951; Brickey and Gorham 1989; Bristowe 1932; Conconi and Bourges 1977; Gope and Prasad 1983; Kevan 1991; Taylor 1975); East Asia (Bodenheimer 1951; Conconi and Bourges 1977; Hoffman 1947; Mitsuhashi 1988; Pierce 1915).

Central America and the Caribbean. In Mexico, J. R. E. de Conconi and collaborators (1984) have documented the consumption of a very wide variety of insects. They found 101 species of insects regularly consumed, fresh, roasted, or fried, at various stages of development. These belong to 31 families in nine

orders and include dragonflies, grasshoppers, bugs, lice, treehoppers, cicadas, caddis flies, butterflies, moths, flies, ants, bees, and wasps (Conconi et al. 1984).

Entomophagy has been practiced for a long time in Mexico (Conconi and Bourges 1977), where many insect dishes are considered delicacies. A well-known

example is *abauatle*, a mixture of hemiptera eggs. It was first reported in 1649 by Francisco Hernandez and was still consumed in 1933 (Ancona 1933). The eggs, from four species of aquatic hemiptera (two species of *Krizousacoriza* and two of *Corisella*), were collected on specially prepared bundles of leaves left in the lake water for several days during the principal egg-laying season. This mixture of eggs (said to have a shrimp-like flavor) was then fried with chicken eggs. The insect eggs were also dried and used as a condiment in the preparation of a traditional Christmas Eve dish, *revoltijo*.

The adult insects of both *Krizousacoriza* and *Corisella* were also consumed, usually along with a fifth species, *Notonecta unifasciata*. The adults were gathered in baskets, sun-dried, and sold on the street, as *mosco para los pájaros*. They were also ground and made into small cakes that were grilled. Both the eggs and adults of these species of aquatic insects were sold as fish food in Europe and in the United States (Ancona 1933).

South America. The giant queen ants of the genus *Atta* are prized as a gastronomical delicacy in Colombia. *Atta* are leaf-cutting ants considered to be a major pest in some agricultural areas. The queens of the genus are very large ants that swarm in the early rainy season and can be easily collected as they leave the nest. The gathering and toasting of these ants is a prosperous cottage industry in northeastern Colombia, and they are marketed throughout the country. Ant consumption can be traced to precolonial times: Gonzalo Jimenez de Quesada, founder of the Colombian capital, Santa Fe de Bogotá, described their use by local peoples in the highlands in 1555 (*Nación* 1989).

The consumption of a wide variety of insects (representing 10 orders and more than 50 different species) has been reported among Amerindian groups in South American rain forests. The insects that appear to be most commonly consumed are ants of the genus *Atta*, palm grubs, and caterpillars of various sorts. The consumption of the queen ants occurs throughout the Amazon Basin (Wallace 1853; Bodenheimer 1951; Denevan 1971; Weber 1972; Ruddle 1973; Dufour 1987). The first description of their use was provided by the naturalist Alfred Wallace (1853: 242-3):

The part eaten is the abdomen, which is very rich and fatty from the mass of undeveloped eggs. They are eaten alive; the insect being held by the head as we hold a strawberry by its stalk, and the abdomen being bitten off, the body, wings and legs are thrown down to the floor, where they continue to crawl along apparently unaware of the loss of their posterior extremities.

Palm grubs, or palm worms, are the large, fatty, legless larvae of wood-boring weevils (genus *Rhyncophorus*) found in the pith of felled palm trees. The weevils are ubiquitous in tropical rain forests, and the

utilization of their larvae as a valued food has been widely reported (Bodenheimer 1951; Chagnon 1968; Clastres 1972; Beckerman 1977; Lizot 1977; Milton 1984; Dufour 1987). E. Bancroft, writing in the eighteenth century, claimed they were also highly esteemed by Europeans in Surinam, particularly by the French, who mixed the roasted grubs with bread crumbs, salt, and pepper (Bancroft 1769, cited in DeFoliart 1990). In northwest Amazonia they are eaten both raw and after smoke-drying over a slow fire, as components of normal meals and as snacks. Amerindians "cultivate" palm grubs in the sense that they fell palm trees with the expectation that they will be invaded by weevils and a crop of grubs will be available for harvest in two to three months (Clastres 1972; Beckerman 1977; Dufour 1987).

The consumption of other types of grubs or larvae is frequently mentioned in ethnographic reports. In the northwest Amazon, the larvae of wood-boring beetles are extracted from felled trees (Dufour 1987). One group of Amerindians in the Peruvian montaña region "cultivate," or at least protect, an unidentified grub called *poshori*, found in piles of maize cobs (Denevan 1971).

Caterpillar eating has been reported for a number of Amerindian groups (Lizot 1977; Dufour 1987), but the actual species eaten have been identified for only one group - the Yukpa (Ruddle 1973). In the northwest Amazon region, five different species of caterpillars have been collected, two of them in large amounts (Dufour 1987). Less commonly ingested are soldier ants, termites, wasp and bee brood, and other larvae (grubs) of various sorts. The consumption of adult beetles seems to be even less common, but it has been reported for the Yukpa (Ruddle 1973) and Ache (Hurtado et al. 1985).

Europe

Although entomophagy is no longer common in Europe, insects were eaten throughout the continent at one time. Cock chafer grubs were consumed by peasants and mountain inhabitants until the 1800s and were an important source of protein in Ireland during the famine of 1688 (Bodenheimer 1951). In earlier times, those who could not afford a diet rich in protein and fat relied on insect sources to supplement their predominantly carbohydrate intake (Bodenheimer 1951).

The Greeks and Romans, who heavily influenced European culture, enjoyed insects as a food source (Bequaert 1921). Ancient Greeks apparently considered grasshoppers a delicacy, and even Aristotle reported eating cicadas. He considered them tastiest just before the final instar, but females, laden with eggs, were deemed delicious as well (Bates 1959). The Greeks and Romans also ate a large Melolonthid grub thought to be *Lucanus cervus*, which, according to Pliny, was fattened before consumption (Bequaert 1921).

Africa

The consumption of insects in Africa is still quite widespread (DeFoliart 1989), with caterpillars among the most popular. A recent survey found that 69 percent of Africans interviewed had eaten one or more species of caterpillar or lived in households where at least one species of caterpillar was consumed (Ashiru 1988). The mopanie worm (*Gonimbrasia belina*), the so-called snack that crawls, is one of the best-known edible caterpillars. It is eaten fried, dried, stewed in tomato sauce, and even raw (Brandon 1987). Mompani worms appear to constitute an important source of protein for rural dwellers (Ivbi-jaro 1990) and were reportedly preferred to meat in some areas when they were in season (Quin 1959).

Caterpillars have become something of a trend on menus around Africa and are currently marketed by a firm in Johannesburg (Brandon 1987; *Zimbabwe Herald* 1988). Because they occur in large aggregations, the collection of caterpillars is easier than that of some other insects (Chavanduka 1975). In fact, collection rates of 40 pounds per day have been reported (Brandon 1987).

Locusts also play a significant role in the diet of Africans, particularly *Schistocerca gregaria*, the desert locust (Quin 1959; DeFoliart 1989). Indeed, in the not-so-distant past, the locusts have been so popular that locust swarms were welcomed and their arrival attributed to a mighty spirit (Bequaert 1921). The female locusts loaded with eggs were most commonly eaten, and often a soup was made of the eggs (Berensberg 1907; Bequaert 1921). High in protein and fat, locusts may be an excellent supplement to local diets that consist mainly of carbohydrates (Chavanduka 1975). P. J. Quin (1959) also notes that locusts played a significant role in sustenance and were the most popular meat of all among the Pedi of the Transvaal.

Termites are also utilized as food in Africa, especially during the early rainy season when the reproductive forms fly from the nest in large swarms and can be collected (Brickey and Gorham 1989; Kumar 1990). At one time, termites were such an important addition to the diet that their mounds were fought over as property (Bequaert 1921). Local people knew when these swarms would occur and they would cover mounds for purposes of collection (Curran 1939). Termites can be eaten raw (with care taken to avoid being bitten by the termite first) (Curran 1939) and fried, and they are also roasted. Termites not only are high in fat and protein but also contain a significant amount of lysine, which is not found in maize, a main constituent of the African diet (Chavanduka 1975; Oliveira et al. 1976).

Palm grubs, such as those of the genus *Rynchophorus*, are still consumed in Cameroon, Angola, and other parts of Africa and are occasionally raised specifically for this purpose (Oliveira et al. 1976; DeFoliart 1990). They may be prepared by stewing, frying, or grilling, but one creative recipe suggests boiling them inside coconuts instead (Grimaldi and Bikic 1985, cited in DeFoliart 1990).

Although a recent study indicates that 10 percent of the animal protein in the diet of local people in Zaire comes from insect foods, the general trend in Africa is toward a reduction in entomophagy (DeFoliart 1990; Kumar 1990). This development is, perhaps, due to an increased use of pesticides as well as to exploitation of new habitats in order to accommodate growing populations, both of which have reduced the number of insects to which populations have access (Chavanduka 1975).

Asia

Middle East and South Asia. With the exception of locusts, whose consumption continued into recent times, the Middle East and South Asia have not had a strong history of entomophagy (Bodenheimer 1951). The desert locust, *S. gregaria*, was a major source of food in the Middle East, with up to 9,000 kilograms gathered per day in some areas (DeFoliart 1989). These locusts were prepared by cooking in salt water and then sun-drying, after which they were traded at the market (DeFoliart 1989). During times of famine locusts were frequently ground into flour from which fried cakes were made (Bodenheimer 1951). In 1988, Saudi Arabians were also observed eating locusts, grilled like shrimp, after the worst locust plague in 25 years descended on the west coast (San Francisco *Chronicle* 1988).

Locusts were especially prevalent in the diet of nomadic people such as the Bedouins, who welcomed the periodic swarms (Bodenheimer 1951). The best-known incident in that part of the world involving locust eating, however, concerns John the Baptist's limited fare of locusts (St. John's bread) and honey during his ordeal in the desert. He was observing the decree of Moses found in Leviticus 9:22: "These ye may eat; the locust after his kind and the bald locust after his kind, and the cricket after his kind and the grasshopper after his kind."

A recent study of three tribes in Assam, India, has revealed a surprising number of locusts eaten on a mass scale (Bhattacharjee 1990). No other contemporary insect consumption is documented for South Asia except in Nepal. The Nepalese in the Himalayan foothills are said to prepare a native dish, called *bakuti*, in which the larvae and pupae of giant honeybees are extracted from the wax comb and cooked until they resemble scrambled eggs (Burgett 1990).

Southeast Asia. Entomophagy is still relatively popular in Southeast Asia. In fact, Thailand currently exports frozen steamed ant larvae and pupae to the United States as specialty foodstuffs (Brickey and Gorham 1989). In North Thailand, bee brood is prepared for consumption by steaming entire honeycombs wrapped in banana leaves (Burgett 1990). Such grubs and pupae must be an important food among local people because, as F. S. Bodenheimer (1951) notes, the collection of combs is done at great risk of being stung.

The consumption of beetles has been especially popular throughout Southeast Asia, with 24 species in six families listed by Bodenheimer (1951) alone. He reported that Dytiscid beetles, such as *Cybister* and Hydrophilid beetles, seem to have been the most frequently eaten and those most commonly sold at market (Hoffman 1947). The larvae of longicorn beetles and weevils were also sought by the people of Laos (Bodenheimer 1951). The larvae of *Rhynchophorus schab*, which are found in coconut palms, were roasted, and other *Rhynchophorus* larvae were fattened and sold for a good price at the market.

While in Laos and Siam (now Thailand) in the 1930s, W. S. Bristowe (1932) also noted the popularity of larval as well as adult beetles in this region. Among those he cited as favorites were dung beetles (Scarabaeidae), which were served curried, and the giant water beetle (*Lethocerus indicus*), a great delicacy that was said to taste like Gorgonzola cheese. Bristowe also observed dishes of water, set out with lighted candles in the center, to attract dragonflies and termites. These insects would singe their wings on the candles and fall into the water to be subsequently collected.

Bristowe (1932) claimed that the local people's knowledge of the life history of these food insects was very thorough and that they truly enjoyed eating the insects and were not consuming them strictly out of necessity. Furthermore, he hypothesized that by eating pest insects the people were reducing the amount of damage done to their crops.

The green weaver ant (*Oecophylla smaragdina*) was used throughout South and Southeast Asia in a condiment as well as a drink; both were said to have a rather acidic flavor (Bequaert 1921).

East Asia. Insect consumption is still quite common in some parts of East Asia. Perhaps the most well-known insect eaten in this region is the pupae of the silkworm *Bombyx mori*. These pupae are exported from Korea to the United States (Brickey and Gorham 1989) and are also eaten in China, according to a recent study done by the Institute of Insect Research in Yunnan Province, China (The Research Institute 1990). W. E. Hoffman (1947) observed that the pupae were sold throughout the silk district and often cooked immediately after the silk was unraveled, so as not to waste the valuable resource that accompanied the silk.

The larvae and pupae of Hymenopterans are another food resource currently used in East Asia. Five species of immature bees and wasps utilized as food were identified by the Institute of Insect Research in Yunnan Province. B. Hocking and F. Matsumura noted in 1960 that Japan had been exporting bee pupae in soy to the United States. Interestingly, one of the favorite dishes of Japan's late Emperor Hirohito was wasp pupae and larvae over rice (Mitsuhashi 1988). Bodenheimer (1951) also documented the consumption of the larvae and pupae of *Vespa*, a wasp, in Japan.

Locusts are presently consumed in East Asia, partic-

ularly those of the genera *Oxya* and *Locusta* (The Research Institute 1990b). Bodenheimer (1951) described the sale of grasshoppers fried in sesame oil in the local markets in China earlier in this century. Other insects consumed in East Asia included *L. indicus*, the giant water beetle, and caterpillars infected with *Cordyceps*, a fungus. These caterpillars were tied in bundles by the fungus filaments and sold as a delicacy (Hoffman 1947).

Pacific Islands. According to V. B. Meyer-Rochow (1973), insect eating is still a part of the native cultures in Papua New Guinea. Most widely documented are the sago palm grub feasts in which hundreds of pounds of *Rhynchophorus bilineatus* larvae are wrapped in banana leaves and roasted. In fact, three species of palm grubs are exploited regularly, along with other larvae and adults of large beetles including many of the family Scarabaeidae (Meyer-Rochow 1973).

New Guineans also consume a great number of Orthopterans, including two species of stick insects, two of mantids, and the female of a species of locust (Meyer-Rochow 1973). Meyer-Rochow hypothesized that insect consumption provided needed dietary protein and fat because the greatest variety and quantity of insects were consumed in areas with the highest population density as well as the lowest amounts of available animal protein.

Termites were also very popular in the Pacific Islands, especially the sexual winged forms, which are higher in fat content. The predominant species eaten, *Macrotermes*, were caught after they flew into fires, burning their wings. In the Philippines, natives were said to eat the locusts that consumed their potato vines - a unique form of pest control (Bodenheimer 1951).

Australia and New Zealand

Despite a general reduction in the use of insects as food worldwide, entrepreneurs in Australia have begun to introduce insects to their commercial food market (Putting Insects on the Australian menu 1990). The insects considered to have the best potential for success include black honey ants, witchetty grubs (the larvae of a Cossid moth), bardi grubs (the larvae of a Cerambycid beetle), and *Trigona* bees. All of these insects have been supplied from around Australia where they are local delicacies. Many restaurants in Australia also now include insects on their menus (Putting Insects on the Australian menu 1990).

The black honey ant (*Camponotus inflatus*) is similar to the honey ant found throughout North and Central America. A modified worker ant with an enlarged body the size of a grape and filled with nectar, the black honey ant was highly sought after by Aboriginal Australians (Bourne 1953) and even considered a totem animal by some clans (Flood 1987; Conway 1990). The digging up of these ants is considered an important traditional practice and is still taught to children (Conway 1990). Other species of

ants were also important food resources for Aboriginal Australians (Bodenheimer 1951). Tasmanians had a number of words for the various ants that they identified, and traditional feasts were held during the periods that ant pupae could be obtained (Sutton 1990).

Witchetty grubs were also an important food of Australian Aborigines (Flood 1987). The name "witchetty grub" refers to several types of grubs (Campbell 1926) and probably includes the larvae of the Cossid moth (*Xyleutes leucomochla*), the giant ghost moth (Hepialidae), and the longicorn beetle (Cerambycidae) (Bodenheimer 1951; Sutton 1990).

Among the most unique and well-documented examples of entomophagy in Australia were the annual Bugong moth feasts that occurred in the Victorian Alps until the 1890s. Moths of *Agrotis infusa* migrate from the plains of New South Wales to aestivate (the summer equivalent of hibernation) every year in the rock crevices and caves of the Bugong Mountains, from which their name was derived (Flood 1987). Enormous numbers of moths have been counted (up to 17,000 per square meter) in some of these crevices and found on the floors of caves to depths of over 1 meter (Flood 1980). Many different tribes of Aboriginal Australians traditionally gathered to feast on the moths, and evidence of these feasts has been carbon dated to about 1,000 years ago (Flood 1980). The moths were smoked out by thrusting a burning bush into the cracks of the rock, then they were thrown into hot ashes or sand to remove the wings and legs (Campbell 1926; Flood 1980). These moths are said to have a nutty flavor and are high in both protein and fat (Campbell 1926; Sutton 1990).

Insects Are Good to Eat

General Patterns of Use

The great number and diversity of insect species (over 600) that have been or are used as food are listed in Table II.G.15.1. The largest variety of species appears to be in the tropics, which probably reflects the greater faunal diversity in these areas and the larger size of individual organisms. But it may also reflect to some extent the amount of research done on the topic in such climates. Comparatively fewer species appear to have been consumed in the more temperate regions of North America, Asia, and Europe. Fewer insect species inhabit these areas, and in general, the size of individual organisms is smaller.

A closer look at the insects listed in Table II.G.15.1 reveals that many of the same types of insects (i.e., in the same orders and families) are consumed in different parts of the world. The most widely consumed are probably the locusts (Acrididae), termites (Termitidae), and palm grubs (Curculionidae) (see Table II.G.15.2). These insects share two characteristics that are significant to their use as human food: (1) Individual

insects are relatively large; and (2) individuals are highly aggregated during at least part of their life cycle. Locusts, for example, are among the largest of insects. They are also gregarious and can form swarms of thousands of individuals. Termites are generally smaller in size but are ubiquitous in tropical ecosystems. The species employed as food can be found in large, easily recognized colonies (for example, *Macrotermes* spp. in Africa), and the reproductive forms of some species can be easily captured during mating flights. Palm grubs, the larvae of weevils, can be as big as sausages. Over 50 percent of the Curculionidae reported are from the genus *Rynchophorus*, a genus found in the Americas, Africa, and Asia, which includes some of the largest of all herbivorous beetles (Crowson 1981). In addition, palm grubs can be harvested in sizable quantities from the pith of felled palm trees.

About 50 percent of the insects listed in Table II.G.15.1 are consumed in immature stages of the life cycle. The advantage to human consumers is that the immature stages are soft-bodied, typically high in fat, usually the largest form of the life cycle, and often in the stage of the life cycle when individual insects can be found in the greatest aggregations. In the order Lepidoptera (butterflies and moths), for example, the larval stage is the stage of the life cycle in which the insect is largest and has the highest energy (caloric) content. Compared to the larvae in the last instar, the adult forms have a lower body mass and a hardened exoskeleton, which reduces their digestibility. The larval stage is also often the stage of the life cycle at which the organisms are the most numerous and highly aggregated, whereas the adults are more mobile and widely dispersed. Furthermore, the larval stage is often long, as opposed to the shorter adult stage (Daly 1985), and therefore offers a greater harvesting opportunity.

Table II.G.15.2. *Number of species of the insects most commonly consumed throughout the world by geographic region (percentage consumed in immature forms in parentheses).*

Geographic region	Number of species ^a			
	Loc	Ter	PG	Total
North America	10	0	1	50 (46%)
Central America, Caribbean	16	0	3	117 (81%)
South America	8	5	4	61 (57%)
Europe	2	0	0	20 (54%)
Africa	10	19	6	155 (63%)
Middle East, South Asia	6	2	0	29 (33%)
Southeast Asia	4	1	3	67 (34%)
East Asia	5	0	1	60 (40%)
Pacific Islands	4	6	4	59 (27%)
Australia, New Zealand	2	2	0	46 (68%)
Total	67	35	22	664

^aLoc = locusts, Ter = termites; PG = palm grubs.

Table II.G.15.3. *Nutritional value of some insects consumed in tropical South America compared with other animal foods; composition per 100 grams edible portion.*

Food	Moisture (%)	Energy (kcal)	Protein (g)	Fat (g)
Ants, female sexuals (<i>Atta sexdens</i>)	6.1	628	39.7	34.7
Ants, female sexuals (<i>Atta cephalotes</i>)	6.9	580	48.1	25.8
Termites, soldiers (<i>Syntermes</i> sp.)	10.3	467	58.9	4.9
Palm grubs, smoke-dried (<i>Rhynchophorus</i> sp.)	13.7	661	24.3	55.0
Caterpillars, smoke-dried (various species)	11.6	425	52.6	15.4
River fish, smoke-dried	10.5	312	43.4	7.0
Tapir, smoke-dried	10.3	516	75.4	11.9

Source: Adapted from Dufour (1987).

Food Value of Insects

In terms of nutrition, insects are comparable to other animal foods. The composition of some insects consumed in tropical South America is shown in Table II.G.15.3, and values for two other common sources of animal protein – dried fish and dried meat – are provided for comparison. The energy value of the insects is high, between 425 and 661 kcal per 100 grams, with the energy values for female alate ants (*Atta* sp.) and *Rhynchophorus* larvae the highest, because both of these are rich in fat. The amount of protein in insects is also relatively high, but its quality appears to be somewhat lower than that of vertebrates (Conconi and Bourges 1977). In relation to more familiar foods, the composition of ants, palm grubs, and caterpillars is comparable to goose liver, pork sausage, and beef liver, respectively. The composition of termite soldiers is roughly comparable to that of non-oily fish, although the latter contain more protein.

Unfortunately, the actual nutritional importance of insects in human diets is not well understood because there is little information on quantities actually consumed.

Summary and Conclusions

Insects are ubiquitous and have been included in human diets in most areas of the world at some time. They are still a component of many current diets. The inclusion of insects in these diets should not be regarded as a mere curiosity, nor should their nutritional importance be overlooked because of their small size. Insects are not only a food prized by many peoples but also a good source of energy, animal protein, and fat. Their dietary importance varies from region to region and group to group, but the wide-

spread practice of entomophagy deserves attention in any description or evaluation of human diets.

Darna L. Dufour
Joy B. Sander

Bibliography

- Aldrich, J. M. 1912a. Flies of the Leptid genus *Atherix* used as food by California Indians. *Entomological News* 23: 159–63.
- 1912b. Larvae of a Saturniid moth used as food by California Indians. *Journal of the New York Entomological Society* 20: 28–31.
- Ancona, L. 1932. Los jumiles de Texco. *Anales del Instituto de Biología de Universidad Nacional de México* 3: 149–62.
1933. El ahuate de Texcoco. *Anales del Instituto de Biología de Universidad Nacional de México* 4: 51–69.
- Ashiru, M. O. 1988. The food value of the larvae of *Araphe venata*. *Ecology of Food and Nutrition* 22: 313–20.
- Bates, M. 1959. Insects in the diet. *American Scholar* 29: 43–52.
- Beckerman, S. 1977. The use of palms by the Bari Indians of the Maracaibo Basin. *Principes* 22: 143–54.
- Bequaert, J. 1921. Insects as food. *Natural History* 21: 191–201.
- Berensberg, H. P. von. 1907. Uses of insects as food, delicacies, medicines, or in manufactures. *Natal Agricultural Journal* 10: 757–62.
- Bhattacharjee, P. C. 1990. Food insect consumption in India. *The Food Insects Newsletter* 3: 8.
- Bodenheimer, F. S. 1951. *Insects as human food*. The Hague.
- Bourne, G. H. 1953. The food of the Australian Aboriginal. *Nutrition Society Proceedings* 12: 58–65.
- Brandon, H. 1987. The snack that crawls. *International Wildlife* 17: 16–21.
- Brickey, P. M., and J. R. Gorham. 1989. Preliminary comments on federal regulations pertaining to insects as food. *The Food Insects Newsletter* 2: 1, 7.
- Bristowe, W. S. 1932. Insects and other invertebrates for human consumption in Siam. *Transactions of the Entomological Society of London* 80: 387–404.
- Brues, C. T. 1946. *Insect dietary; an account of the food habits of insects*. Cambridge, Mass.
- Burgett, M. 1990. Bakuti – a Nepalese culinary preparation of giant honey bee brood. *The Food Insects Newsletter* 3: 1.
- Campbell, T. G. 1926. Insect foods of the Aborigines. *The Australian Museum Magazine* 2: 407–10.
- Chagnon, N. 1968. *The fierce people*. New York.
- Chavanduka, D. M. 1975. Insects as a source of protein to the African. *The Rhodesia Science News* 9: 217–20.
- Clastres, P. 1972. The Guyaki. In *Hunters and gatherers today*, ed. M. G. Bicchieri, 138–73. New York.
- Conconi, J. R. E. de, and H. R. Bourges. 1977. Valor nutritivo de ciertos insectos comestibles de México y lista de algunos insectos comestibles del mundo. *Anales del Instituto de Biología, Seria Zoología* 48: 165–86.
- Conconi, J. R. E. de, J. M. P. Moreno, C. M. Mayaudon, and F. R. Valdez. 1984. Protein content of some edible insects in Mexico. *Journal of Ethnobiology* 4: 61–72.
- Conway, J. R. 1990. A program profile: Honey ants and the Australian Aborigines. *The Food Insects Newsletter* 3: 2.

- Crowson, R. A. 1981. *The biology of coleoptera*. New York.
- Curran, C. H. 1939. On eating insects. *Natural History* 43: 84-9.
- Daly, H. V. 1985. Insect morphometrics. *Annual Review of Entomology* 30: 415-38.
- DeFoliart, G. 1989. The human use of insects as food and animal feed. *Bulletin of the Entomological Society of America* 35: 22-35.
1990. Hypothesizing about Palm weevil and Palm Rhinoceros beetle larvae as traditional cuisine, tropical waste recycling, and pest and disease control on coconut and other palms - can they be integrated? *The Food Insects Newsletter* 3: 1-6.
1991. Forest management for the protection of edible caterpillars in Africa. *The Food Insects Newsletter* 4: 1-2.
- Denevan, W. M. 1971. Campa subsistence in the Gran Pajonal, eastern Peru. *Geographical Review* 61: 496-518.
- Dufour, D. L. 1987. Insects as food: A case study from the northwest Amazon. *American Anthropologist* 89: 383-97.
- Eerde, E. 1980-1. Butterflies in your stomach? *R&D Mexico* Dec./Jan., 6-8.
- Engelhardt, G. P. 1924. The Saturniid moth, *Coloradia Pandora*, a menace to pine forests and a source of food to Indians in eastern Oregon. *Bulletin of the Brooklyn Entomological Society* 19: 35-7.
- Fenenga, G. L., and E. M. Fisher. 1978. The Cahuilla use of *Piyatem*, larvae of the White-Lined Sphinx Moth (*Hyles lineata*), as food. *The Journal of California Anthropology* 5: 84-90.
- Fladung, E. B. 1924. Insects as food. *Maryland Academy of Sciences Bulletin*, October: 5-8.
- Flood, J. 1980. *The moth hunters: Aboriginal prehistory of the Australian Alps*. Canberra, Australia.
1987. Moth hunters of the southeastern highland. In *Australians to 1788*, ed. D. J. Mulvaney and J. P. White, 257-91. Sydney.
- Fowler, C., and N. Walter. 1985. Harvesting Pandora moth larvae with the Owens Valley Paiute. *Journal of California and Great Basin Anthropology* 7: 155-65.
- Gope, B., and B. Prasad. 1983. Preliminary observation on the nutritional value of some edible insects of Manipur. *Journal of Advanced Zoology* 4: 55-61.
- Gorham, J. R. 1979. The significance for human health of insects in food. *Annual Review of Entomology* 24: 209-24.
- Harris, W. V. 1971. *Termites: Their recognition and control*. London.
- Hocking, B., and F. Matsumura. 1960. Bee brood as food. *Bee World* 41: 113-20.
- Hoffman, W. E. 1947. Insects as human food. *Proceedings of the Entomological Society of Washington* 49: 233-7.
- Holt, V. M. 1885. *Why not eat insects?* London.
- Hurtado, A. M., K. Hawkes, K. Hill, and H. Kaplan. 1985. Female subsistence strategies among Ache hunter-gatherers of eastern Paraguay. *Human Ecology* 13: 1-28.
- Ivbijaro, M. 1990. Insect consumption in Nigeria. *The Food Insects Newsletter* 3: 4.
- Kapoor, V. C. 1985. *Perspectives in insect systematics*. New Delhi.
- Kevan, D. K. McE. 1991. The eating of stick insects by humans. *The Food Insects Newsletter* 4: 7.
- Kodondi, K. K., M. Leclercq, and F. Gaudin-Harding. 1987. Vitamin estimations of three edible species of Attacidae caterpillars from Zaire. *International Journal for Vitamin and Nutrition Research* 57: 333-4.
- Kumar, S. K. 1990. Insect consumption in Zambia. *The Food Insects Newsletter* 3: 4.
- Lizot, J. 1977. Population, resources and warfare among the Yanomami. *Man* (n.s.) 12: 497-517.
- Madsen, D. B. 1989. A grasshopper in every pot. *Natural History* 7: 22-5.
- Madsen, D. B., and J. E. Kirkman. 1988. Hunting hoppers. *American Antiquity* 53: 593-604.
- Malaisse, F., and G. Parent. 1980. Les chenilles comestibles du Shaba meridional (Zaire). *Naturalistes Belges* 6: 12-24.
- Meyer-Rochow, V. B. 1973. Edible insects in three different ethnic groups of Papua and New Guinea. *The American Journal of Clinical Nutrition* 26: 673-7.
- Milton, K. 1984. The protein and carbohydrate resources of the Maku Indians in northwestern Amazonia. *American Anthropologist* 86: 7-27.
- Mitsuhashi, J. 1988. Rice with cooked wasps: Emperor Hirohito's favorite dish. *The Food Insects Newsletter* 1: 2.
- Nación (Bogota). 1989. June 26.
- Oliveira, S., P. de Carvalho, B. de Sousa, and M. Simao. 1976. The nutritional value of four species of insects consumed in Angola. *Ecology of Food and Nutrition* 5: 91-7.
- Pierce, W. D. 1915. The uses of certain weevils and weevil products in food and medicine. *Proceedings of the Entomological Society of Washington* 17: 151-5.
- Putting insects on the Australian menu. 1990. *The Food Insects Newsletter* 3: 2.
- Quin, P. J. 1959. *Food and feeding habits of the Pedi*. Johannesburg.
- The Research Institute of Insect Resources in Yunnan Province, China. 1990. *The Food Insects Newsletter* 3: 6.
- Ruddle, K. 1973. The human use of insects: Examples from the Yukpa. *Biotropica* 5: 94-101.
- San Francisco Chronicle. 1988. September 22.
- Simmonds, P. L. 1885. *The animal resources of different nations*. London.
- Sutton, M. Q. 1985. The California salmon fly Pteronarcys Californica (Plecoptera) as a possible food source in N. E. California. *Journal of California and Great Basin Anthropology* 7: 176-82.
1988. Insects as food: Aboriginal entomophagy in the Great Basin. *Ballena Press Anthropological Papers*. Menlo Park, Calif.
1990. Aboriginal Tasmanian entomophagy. In *Ethnobiology: Implications and applications*, ed. D. A. Posey, W. L. Overall, C. R. Clement, et al. Proceedings of the First International Congress of Ethnobiology, Belém, Brazil.
- Swezey, S. L. 1978. Barrett's armyworm: A curious ethnographic problem. *The Journal of California Anthropology* 5: 256-62.
- Taylor, R. L. 1975. *Butterflies in my stomach*. Santa Barbara, Calif.
1976. *Entertaining with insects*. Santa Barbara, Calif.
- Turk, D. 1991. Book review: Les insectes comme aliments de l'homme. *The Food Insects Newsletter* 4: 5-8.
- Wallace, A. R. 1853. On the insects used for food by the Indians of the Amazon. *Transactions of the Royal Entomological Society of London* 2: 241-4.
- Weber, N. A. 1972. *Gardening ants: The Attines. Memoirs of the American Philosophical Society*, Vol. 92.
- Yturbide, T. C. 1986. *Presencia de la comida prehispanica*. Mexico City.
- Zimbabwe Herald. 1988. April 12.

II.G.16 🐫 Llamas and Alpacas

The llama (*Lama glama*) and alpaca (*Lama pacos*) are among the few domesticated ungulates whose most important function has not been that of providing food for the people who control them. The llama has been kept primarily as a beast of burden, whereas the more petite alpaca is most valued as a source of an extraordinarily fine fleece. These South American members of the camel family may share a common biological ancestry from the guanaco (*Lama guanicoe*), for although they have long been designated as separate species, the closeness of the relationship is reflected in fertile offspring when they crossbreed. An alternative point of view now gaining in popularity is that the alpaca descended from the vicuña (*Lama vicugna*), since both animals are about the same size (44 to 65 kilograms [kg]) and both have the capacity to regenerate their incisor teeth. The distribution of both the llama and the alpaca has been traditionally centered in the Andean Highlands of Peru and Bolivia, with peripheral populations of the former in Chile, Argentina, and Ecuador. In the past three decades growing interest has increased their population on other continents, especially in North America.

Camelid Meat as Human Food

Both animals have been an important source of food in the part of the central Andes where husbandry has been most intensive. In neither case, however, are they raised primarily for their flesh, which is consumed after their most valued functions diminish with age. However, llamas possibly had a more important meat func-

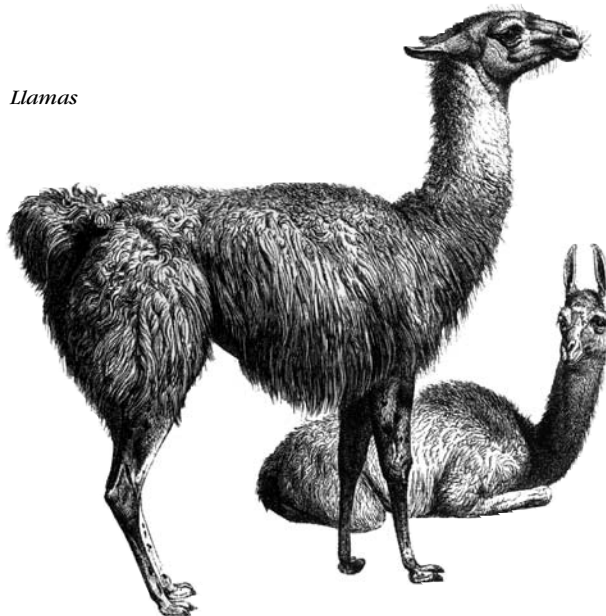
tion in the pre-Pizarro Andes before the introduction of European barnyard creatures. The movement of herds from the highlands to the coast could have been a way both to transport goods and to move protein-on-the-hoof to the more densely populated coast, where at the time, meat was much rarer than in the highlands (Cobo 1956). David Browman (1989) suggested that when camelid utilization in the highlands expanded northward, starting around 1000 B.C., and long before the Inca civilization was established, meat production appeared to have been the most important use of these animals. But whether of primary or secondary importance, the protein and fat supplied by this meat have contributed to the health of the animals' Andean keepers, whose diet consists mainly of starch.

Some fresh meat is consumed soon after an animal has been butchered. Recently butchered muscle meat is normally whitish or pinkish. Travelers' reactions to llama flesh have been mixed, although it is not always clear that the opinion is based on actual consumption. J. J. Von Tschudi (1848), for example, described it as "spongy in texture and not agreeable in flavor." It is reportedly as high in protein (more than 20 percent) as other kinds of meat but lower in fat than mutton, pork, or beef (Bustinza Choque 1984). Sausages and a kind of Andean haggis are made from the blood and intestines (Flannery, Marcus, and Reynolds 1989). The soft parts of the animals may also be fried (*chicharrones*).

Most llama and alpaca meat is dried, which in this region of the world has been the only feasible way to preserve and market it outside the local community. Chunks of the carcass are separated from the bone and cut into thin slices less than 1 centimeter thick. These are usually salted by soaking in a brine solution for several days. Then the meat is exposed for two or three weeks in May and June when night temperatures fall to minus 15° Celsius (C). The alternation of freezing cold at night and intense sunlight in the day dries the meat. Called *charqui* in Spanish (from *ch'arki* in Quechua), the dried meat keeps indefinitely. The term *charqui* has given rise to the English word "jerky," which means strips of dried meat, usually beef. In addition to its transportability, *charqui* is reportedly free of a sarcoporiidian parasite that is commonly found in fresh meat (Calle Escobar 1984: 167).

The main consumers of *charqui* are Indians, who barter for it if they do not produce it themselves. *Charqui* continues to be carried in llama trains to lower ecological zones where it is exchanged for maize, fruit, and coca. It is also seen in the periodic markets on the Altiplano of southern Peru and Bolivia and is a staple food in several mining towns. Its main competitor is dried mutton. After sheep were introduced to the Andes in the sixteenth century and Indians subsequently adopted them as their own, the process for making *charqui* was applied to mutton to make *chalona*. Whether fresh or dried, the meat of llamas or alpacas is stigmatized as Indian food, and many mestizos avoid it

Llamas



simply for that reason. Because these animals are not inspected in certified slaughterhouses, questions have arisen about the wholesomeness of the meat.

Llamas and alpacas are led to slaughter typically when their primary functions are deteriorating. However, some other criteria are considered depending on the place or situation. George Miller (1977) pointed out that the most disobedient llamas of the herd are sometimes the ones chosen to be slaughtered. If carried out as a general practice, this kind of culling acts as a selection mechanism for docility. Llamas may also be culled for slaughter if they are unusually small or weak, qualities that inhibit their function as pack animals. Moreover, alpacas may be culled for slaughter if their fleece is multicolored; the international wool trade much prefers white or light fawn-colored fleece and pays higher prices for it. Hybrids (wari or huarizo) between an alpaca and llama have neither of the advantages of the parents and may be preferentially kept as a meat source.

Butchering may also be the result of a need or desire for ritual products from the animal. Bezoar stones, for example, which sometimes occur in the stomach and intestines of llamas and alpacas, were used in native ritual. They were also valued by the Spaniards, who believed them to be an antidote for poison. Fat (*pichuuvira*), especially of llamas, has uses in rites that combine a number of items, including coca leaves, cigarettes, and wine or spirits. Fetuses of pregnant llamas and alpacas are also prized as a ritual item, especially as a substitute for sacrificing live animals in towns and cities. In La Paz, Bolivia, the native herb market sells dried fetuses; typically they are buried under the doorway of a private home to ensure fertility. Formerly, the hides of llamas were used for making sandals, but since the 1950s, Indian footwear in the Central Andes is made with rubber from worn-out truck tires.

No reliable production figures for camelid meat are available; however, the yearly slaughter in Peru and Bolivia may reach 10 percent of the combined llama and alpaca population of more than 6 million animals, which is more or less divided equally between the two (Sumar 1988: 26). Almost 89 percent of alpacas in the world are in Peru; Bolivia has 57 percent of the world's llamas (Bustanza Choque 1984). On a household basis, a poor family may butcher 4 animals a year, whereas a more affluent herder may process up to 12 head a year (Flores Ochoa 1968). Aside from slaughter, llamas may be eaten after they have died of other causes (a newborn killed by diarrhea, for example) (Novoa and Wheeler 1984: 124).

The zone of intensive husbandry today is confined to elevated areas above 3,900 meters (m), all within a 300-kilometer (km) radius of Lake Titicaca. Not surprisingly, Puno produces much more camelid meat than does any other Peruvian department. The *paramos* (flatlands) of northern Peru, Ecuador, and southern Colombia have rarely contained llamas or alpacas in any sizable numbers, and in much of Colombia and

Venezuela there are none at all. In the llama heartland, the colonial period's native beast of burden was displaced to some degree when the llamas' *trochas* (trails) were widened and mule driving was introduced. When vehicle roads were built in the twentieth century, the displacement was accelerated and the llama population dropped sharply. But that same decline in numbers has not happened with alpacas, which in fact are seen by non-Indians as much more valuable animals.

The Nonmilking Enigma

The nutriment that llamas and alpacas provide do not include milk, and the failure of humans to milk these domesticated herd animals raises some important questions. Unfortunately, written sources have rather frequently contained misinformation, sometimes based on an Old World conceit about the cultural history of the Americas. Examples include assertions that llamas were used as mounts and to pull plows, and that alpacas have been beasts of burden. Encyclopedias, several geography textbooks, articles in *Nature* and *National Geographic*, and even a dictionary of zoology contain this error (Gade 1969).

Similarly, milk has been claimed as one of the useful products of the llama and alpaca, yet scholars who have studied Andean camelid domestication, ecology, and economy are silent on the subject.¹ Analyses of the archaeological context by C. Novoa and Jane Wheeler (1984) as well as Elizabeth Wing (1975) and David Browman (1974) contain no hints that Andean people kept any of the camelids for their milk. The Spanish chroniclers who wrote of the Inca say nothing about the native people using milk in any form from either species. Quechua, the language of the Inca and their descendants, differentiates human milk (*ñujñu*) and animal milk (*wilali*), but that lexical specificity by itself says nothing about any possible human use pattern of the latter.

If domestication is viewed as a process rather than a sudden occurrence, one might suppose that llamas and alpacas were developed for their milk. If we accept the reports of archaeologists who specialize in camelid bones, then it appears that domestication of the llama and alpaca had occurred by 4000 B.C. (Browman 1989). Thus, one or both of these camelids have been manipulated by humans for some 6,000 years, which (if time were of the essence in domestication) should have turned the animals into abundant milk producers.

If, however, a high level of civilization is critical in explaining the milking phenomenon, then one might expect that the civilization of the Inca (A.D. 1100–1532) would have developed this area of subsistence activity. The Inca knew enough about heredity to breed white llamas for state sacrifice and to have produced a plethora of food surpluses that testified to their skills as sedentary agriculturists and herders. The Inca and the various cultures that preceded them by a millennium were all sufficiently talented to have

understood the enormous nutritional benefits of milk from their two camelids.

Milking would also have enabled a more intensive use of the grassy puna (highlands) and *paramo* above the altitudes permitting agriculture. As with cattle in the Old World, the milking (and bloodletting) of llamas might have heightened the value of pastoralism and thus encouraged a nomadic way of life. Indeed, the availability of milk would have given the Indians a sort of nutritional freedom to roam the high country above 4,000 m (3,500 m in Ecuador) at will, in search of good pasture for large herds. Certainly such activity would have been more facultative than the obligate relationship with agriculturalists that evolved in the Andes.

Llamas and alpacas have some habits in common that would theoretically predispose them to being milked by their human keepers. Both have a natural tendency to group in corrals at night. Their parturition occurs during the rainy period from January to March when abundant grass is available and the creamy surpluses of lactating females could have been removed with little deprivation for their nursing young. Finally, their general docility resembles that of sheep much more than cattle.

Why the Milking Void?

The aboriginal absence of milking in the Andes elucidates certain cultural-historical questions about domestication as a process. Least persuasive is the argument that the quantities of milk were too small to make milking a worthwhile effort. The old saying "there is no excess" (e.g., insufficient liquid left over for humans once the young have suckled) could have been originally said of cattle, sheep, goats, or water buffalo. Early in her domestication some 7,000 years ago, a cow must have yielded only small amounts of milk to the person willing to extract it. The large milk volume associated with purebred dairy cows today does not predate the nineteenth century.

Indeed, in the face of a minuscule output, the key question is why peoples in the Old World started to milk in the first place. A German cultural geographer, Eduard Hahn (1896), had perhaps the most compelling insight about the prehistoric origins of milking. He believed that milking originally had a religious, not an economic, motive. Bowls of milk were offered to the lunar mother goddess, a cult that regulated spiritual life in ancient Mesopotamia where cattle were first domesticated. The frothy white substance that sustains early life had considerable symbolic power and thus ritualistic value. The quantity of milk extracted was incidental because it had not yet spread from a ritual use among the priestly castes to become nourishment for common people. Nonetheless, expansion of milk supplies to meet the cultic requirements gave an impetus to select those cows found to have large udder capacity and tractability to human interlopers. Slowly, over many generations of selective breeding, milk quantities

increased; one result of this was that the original religious motivation for milking was forgotten, and milk became a common nutriment. In other words, the nutritional benefits of milk became recognized as a sufficient reason to carry on the custom.

Such a process was probably "invented" in one place – the Fertile Crescent of southwestern Asia some 6,000 years ago – and then diffused elsewhere. As this diffusion occurred, milking spread to other domesticated mammals. For example, camels, which replaced cattle in desert environments, were first milked by using the model of bovine milking that had begun two millennia earlier. In fact, Carl Sauer (1952) reasoned that because all the Old World herd animals have been milked, milking may well have been the general motive for domesticating them in the first place.

Quantity of milk aside, another reason put forward for not milking llamas and alpacas has to do with behavioral peculiarities of the species. This line of reasoning posits that a lactating female would not allow people to manipulate the vulnerable underpart of her body, as a defense mechanism to ensure sustenance to the offspring. A related assertion is that llamas and alpacas simply cannot be milked because certain pituitary hormones inhibit the lactating female of these species from releasing her colostrum or milk to any but her own offspring. Although these two characteristics may exist, they nevertheless do not explain the phenomenon of nonmilking, again because of the vast behavioral changes that all dairy mammals have undergone to eliminate their natural refusal strategies. At some point in their domestication, individual animals that did not respond to human intervention were likely to have been culled from the breeding pool. Subsequent generations of the selected species gradually evolved in directions more conducive to human manipulation.

The physiological intolerance of Andean people to milk (lactose intolerance) is still another explanation proposed for the historic nonmilking of llamas and alpacas. As in much of Asia and Africa, most of the native folk of the Americas cannot properly digest fresh milk after they have been weaned due to insufficient amounts of the lactase enzyme in the intestine to break down milk sugars (lactose) (Simoons 1973). When pressed to drink milk, Amerindians often suffer nausea and diarrhea. In Peru, for example, nutritional studies have reported that highland children experience an adverse reaction to milk (Calderón-Viacava, Cazorla-Talleri, and León-Barua 1971; Figueroa et al. 1971).

Yet, as with the arguments already examined, lactose intolerance is still not a convincing explanation for the absence of milking in the Andes. After the Spaniards brought cows, goats, and sheep to them, the Andean peoples followed the European example and began to milk these animals. They used the milk for infant feeding and cheese making. Nursing Indian babies can be fed fresh animal milk with impunity because they still have a high level of the lactase

enzyme, whereas cheese is a low-lactose dairy product that many lactose malabsorbers can consume without ill effects. Cheese has another advantage over liquid milk: It is preservable without refrigeration and thus can be transported long distances. Cows' milk predominates in making cheese, but in dry areas goats' milk is also used. Cheese making from ewes' milk has been an activity in the Altiplano of Bolivia south of Oruro since the colonial period.

In summary, then, the failure to have milked llamas or alpacas was probably not due to their low milk production, nor to the innate behavior of this camelid genus, nor can it be attributed to the lactose intolerance of the people who have kept them. This leaves a disarmingly simple reason for not milking llamas and alpacas. It never occurred to the indigenes to do so. For those who appreciate milk, yoghurt, ice cream, butter, and cheese in their diet, this failure to grasp the possibility of milking may seem illogical. Yet the pre-Columbian civilizations of the New World did not employ several other basic traits of material and nonmaterial culture either, including the concept of the wheel, the arch, and any form of real writing. To this lack of invention must be added the failure of diffusion. Isolated by oceans, the Andean people, prior to conquest, had no opportunity to learn about the idea of milking from European, African, or Asian peoples.

Moreover, the use of llama or alpaca milk for ritual purposes never emerged in the Andes as it did in the Fertile Crescent, although Inca state religion manifested many associations with flocks in its spiritual rhetoric and sacrificial practices (Brotherston 1989). The elaborate rituals concerned with fertility, curing, and divination that persist today in the high Andes do not involve milk in any way. It was camelid fat – easier to transport and less perishable than milk – that acquired the symbolic role in human needs to communicate with the supernatural forces.

As noted, Andean peoples first learned about milking cows and ewes from the Spaniards. It is puzzling, however, that after the introduction of milking, the practice was not subsequently extended to llamas and alpacas, particularly in zones above 4,200 m where they had no competition and where a kind of transhumance has prevailed. It is the case that the transfer of the idea of milking from one mammalian species to another has occurred within historic times. In northern Scandinavia, for example, after observing Swedish settlers milk their cows, the Lapps started to milk the domesticated mammal under their control, the reindeer. But although milking a lactating llama or alpaca might well have been tried on occasion over the past half millennium, it was not a practice that emerged as a use pattern or one that disseminated elsewhere. If milking these animals had been anything more than episodic, it would have been recorded in some travelers' accounts and eventually in the ethnographic record.

Llamas and Alpacas as Future Dairy Animals

The milk of the Andean camelids has received little research attention. One study in Peru, which derived 256 milk samples from 71 different lactating alpacas, found that yields in a 12-hour period varied widely from only 15 to 20 cubic centimeters (cc) of milk in some to as much as 500 cc in others, with 300 cc as a rough average (Moro 1954). Butterfat content was found to be between 3 and 4 percent and the pH ranged from 6.4 to 6.8. Porcelain white in color and with no distinctive odor, the alpaca milk examined in this study was somewhat sweeter and more viscous than cows' milk. These organoleptic characteristics contrast with the unusual flavor and salty taste reported for the milk of an Old World camelid, the dromedary, whose milk is nevertheless much appreciated in parts of the Middle East. More detailed biochemical data on camelid milk puts lipid content at 4.716 plus or minus 1.307 percent (Fernandez and Oliver 1988: 301). Development of either the llama or alpaca as a dairy animal would first entail the considerable work of selecting and breeding animals for higher milk yields.

Wider commercialization of the llama or alpaca as a source of food deserves consideration. They have potential as meat animals because their growth is fast: 9 kg at birth, they can weigh 29 kg at 9 months and 54 kg at 3 years (Bustanza Choque 1984). Carcass yield is reportedly higher than that of sheep or cattle. Although they can thrive on forage that extends well beyond the native Andean puna grasses, any serious herd investment would require better data on feeding requirements. Moreover, the prejudice against both meat and milk from these animals would need to be addressed in order to find profitable ways to market them to the public. North America, which had an estimated 23,000 llamas and alpacas in 1990, is now the center for stock improvement and creative thinking on the considerable potential of these animals. However, both Peru and Bolivia, where the diversity of the llama and alpaca gene pool is greatest, have banned the export of live animals beyond their borders.

Daniel W. Gade

Note

1. One dissenter to the otherwise solid generalization that Andean people do not milk camelids was William E. Carter, a trained anthropologist who died in 1983. Carter (1975) affirmed that llamas were milked, which prompted me to correspond with him about it. In his response dated February 28, 1975, Carter wrote, "I have on numerous occasion seen people making cheese out of llama milk and even eaten it. . . . Not much milk is obtained, of course, nor does the milking occur on a regular basis, but milking of the animal is done." Carter, however, did not state the place or places where he purportedly saw this llama cheese, or if he actually watched those llamas being milked. Could these have been accidental cases of

milk fortuitously salvaged from a lactating female that had just lost her young? None of the scholars who has intensively studied llamas or alpacas in their contemporary Andean setting alluded to them as even occasional sources of milk to their human owners (Flores Ochoa 1968; Webster 1973; Orlove 1977; Flannery et al. 1989).

William Walton (1811: 36), one of the earliest serious non-Spanish observers of llamas, stated that "her milk is scanty, and is never used by the Indians for any purpose, they prefer leaving it to the 'llamitos' which are not weaned before the age of six months."

Bibliography

- Brotherston, Gordon. 1989. Andean pastoralism and Inca ideology. In *The walking larder*, ed. J. Clutton-Brock, 240–55. Boston.
- Browman, David L. 1974. Pastoral nomadism in the Andes. *Current Anthropology* 15: 188–96.
1989. Origins and development of Andean pastoralism: An overview of the past 6,000 years. In *The walking larder*, ed. J. Clutton-Brock, 256–68. Boston.
- Bustanza Choque, A. Victor. 1984. The Camelidae of South America. In *The camelid: An all-purpose animal*, 2 vols., ed. W. Ross Cockrill, 112–43. Uppsala, Sweden.
- Calderón-Viacava, L., A. Cazorla-Talleri, and R. León-Barua. 1971. Incidencia de malabsorción de lactosa en jóvenes Peruanos sanos. *Acta Gastroenterologica Latinamericana* 3: 11.
- Calle Escobar, Rigoberto. 1984. *Animal breeding and production of American camelids*. Lima.
- Cardozo, G. Armando. 1954. *Los auquenidos*. La Paz.
- Carter, William E. 1975. *Bolivia: A profile*. New York.
- Cobo, Bernabe. [1890–93] 1956. *Historia del nuevo mundo*. 2 vols. Madrid.
- Fernandez, F. M., and G. Oliver. 1988. Proteins present in llama milk. I. Quantitative aspects and general characteristics. *Milchwissenschaft* 43: 299–302.
- Figueroa, Rolando B., E. Melgar, N. Jo, and O. L. Garcia. 1971. Intestinal lactase deficiency in an apparently normal Peruvian population. *American Journal of Digestive Diseases* 16: 881–9.
- Flannery, Kent V., Joyce Marcus, and Robert G. Reynolds. 1989. *The flocks of the Wamani: A study of llama herders on the punas of Ayacucho, Peru*. San Diego.
- Flores Ochoa, Jorge A. 1968. *Los pastores de Paratia*. Mexico City.
- Gade, Daniel W. 1969. The llama, alpaca and vicuna: Fact vs. fiction. *Journal of Geography* 68: 339–43.
- Hahn, Eduard. 1896. *Die Haustiere und ihre Beziehungen zur Wirtschaft des Menschen*. Leipzig.
- Miller, George R. 1977. Sacrificio y benéfico de camelidos en el sur del Peru. In *Pastores de puna: Uywamichiq puna-runakuna*, ed. Jorge Flores Ochoa, 193–210. Lima.
- Moro, S. Manuel. 1952–6. Contribución al estudio de la leche de las alpacas. *Revista de la Facultad de Medicina Veterinaria* 7–11: 117–41.
- Novoa, C., and Jane C. Wheeler. 1984. Lama and alpaca. In *Evolution of domesticated animals*, ed. Ian L. Mason, 116–28. New York.
- Orlove, Benjamin. 1977. *Alpacas, sheep, and men*. New York.
- Sauer, Carl O. 1952. *Agricultural origins and dispersals*. New York.
- Simoons, Frederick J. 1973. New light on ethnic differences in adult lactose intolerance. *American Journal of Digestive Diseases* 18: 596–611.
- Sumar, Julio. 1988. Present and potential role of South American camelids in the high Andes. *Outlook on Agriculture* 17: 23–9.
- Von Tschudi, J. J. 1848. *Travels in Peru during the years 1838–1842*. New York.
- Walton, William. 1811. *An historical and descriptive account of the four species of Peruvian sheep called "carneros de la tierra"*. London.
- Webster, Steven. 1973. Pastoralism in the South Andes. *Ethnology* 2: 115–34.
- Wing, Elizabeth. 1975. Hunting and herding in the Peruvian Andes. In *Archaeozoological studies*, ed. A. T. Clason, 302–8. Amsterdam.

II.G.17 Muscovy Ducks

Of the two species of domesticated anatines, the Muscovy duck (*Cairina moschata*) is larger, less vocal, and characterized by a fleshy protuberance on the head of the male. It is a duck of tropical American origin, whose wild ancestors nested in trees, whereas the common duck (*Anas platyrhynchos*) was domesticated in the Old World from the ground-dwelling mallard. The two species can mate and produce offspring, but such offspring cannot reproduce.

The Muscovy duck is misnamed, for it never had any special association with Moscow. Most likely the name is the result of a garbled corruption of *canard musqué* ("musk duck"); however, this French term is not an accurate descriptor either. Depending on area, various Spanish names are used for the duck in Latin America. Among these are *pato criollo* ("native duck"), *pato real* ("royal duck"), *pato almislado* ("musk duck"), *pato machacón* ("insistent duck"), and *pato perulero* ("Peru duck"). In Brazilian Portuguese, it is most commonly called *pato do mato* ("forest duck"). Indigenous names for this bird indicate its New World origin, including *ñuñuma* in Quechua, *sumne* in Chibcha, and *tlalalacatl* in the Nahuatl language of Mexico.

Before the European conquest of the Americas, the bird's apparent distribution extended from north central Mexico to the Rio de la Plata in Argentina (Donkin 1989). It was and still is kept in a wide range of environments, which include the islands of the Caribbean, deserts, humid tropics, temperate plains, and the high elevations of the Andes. Although several colonial chronicles refer to *C. moschata*, such sources do not indicate that these domesticated birds were particularly important to household economies. In Mexico, the turkey has had greater importance in the houseyard. In South America, the Muscovy had no poultry competitors until the Spaniards and Portuguese brought chickens, which – because of their greater egg-laying capacity – were widely adopted.

Precisely how, where, and why the Muscovy duck was domesticated is impossible to say. Two scenarios can be constructed from knowledge of the wild species. One is that the wild duck was trapped – captured alive in fields of crops that it was feeding on. After this it was tamed and began to reproduce in captivity. Alternatively, humans removed fertilized eggs from the nesting holes of wild birds and incubated them until they hatched. The ducklings, which emerge from the egg in 30 days, then imprinted on humans as their surrogate parents. Because the behavioral characteristics of wild Muscovies draw them into ready association with people, their taming and eventual domestication could have been quite common (Nogueira-Neto 1973).

The habitat preference of wild Muscovies, which roost in trees and nest in tree holes, is near rivers and swamps backed by forests. The Caribbean lowlands of Colombia and Panama offer such habitats and today contain wild Muscovies. Yet the just mentioned behavioral characteristics that draw them into association with people make plausible the suggestion that they were domesticated from the wild in several different South and Central American locales as well as the Caribbean lowlands (Nogueira-Neto 1973).

Archaeological evidence charting the past of the Muscovy duck has been derived from bones and from artistic renderings. Osteological material dated as early as 500 B.C. comes from the coastal region of Ecuador. Close by, in northern Peru, potters of the Mochica culture (A.D. 200–700) modeled the Muscovy duck in clay. Abundant zoomorphic evidence of the Muscovy is also found on pottery vessels, gold objects, and stylized masks from pre-Columbian Mexico (Whitley 1973). In view of the latter, a religious role for the bird might be postulated.

Although the domesticated duck has the same species binomial as its wary wild cousin, certain features distinguish *Cairina moschata domestica* from the wild *C. moschata*. Many domesticated Muscovies, for example, have predominantly white plumage – a partial albinism resulting from artificial selection of a sort found in many other domesticated animals. However, feather colors and plumage patterns can vary widely and often resemble those of the wild bird. Truly domesticated birds are unable to fly; artificial selection has reduced their wings to weak appendages. Moreover, the domesticated duck can survive without access to water for swimming or feeding.

Three characteristics of the domesticated drake are his size (more than twice as large as the female), his hisslike vocalization, and his polygamous appetite for constant copulation. In the sixteenth and seventeenth centuries, the odor of musk was also said to characterize this bird, but it is still not clear if such an opinion derived from a distinctive smell of the species, a secondary sexual characteristic of the adult drake, or a dietary element that imparted a peculiar odor to the birds.

In tropical America, Muscovy ducks have been kept for their meat, their fat, and – to a much lesser degree – their eggs. However, they are slaughtered mainly for special occasions, and except where they are kept in large numbers, their role as a food source is peripheral. Indeed, in most parts of tropical America, they do not seem to have a well-defined place in regional cooking, although live birds and their eggs may be sold in some peasant markets. In times of food surpluses, they can be fed and kept alive as a kind of “bank” to be cashed in when they are sold or traded at peasant markets. Normally, it is at lower elevations that farmers raise Muscovies in sizable flocks; in the high Andes, small farmers may keep only two or three.

Raymond Gilmore (1963: 462) has referred to the production of an aromatic powder made from dried duck meat, which counts as another use, as does the large number of insects that Muscovies eat, which extends their usefulness beyond that of a food source. But most of the food Muscovies eat is from plant sources, and thus they are also employed to help keep irrigation ditches clean. On the desert coast of Peru, they are often seen in such ditches – a niche that may have made them initially useful to pre-Columbian irrigation farmers.

In the sixteenth century, the bird was taken to Europe, where its historic pathway of diffusion became confused in the written record, which often did not make the necessary distinction between the imported Muscovy and the common duck already present. In German-speaking Europe, the Muscovy became known as the *Türkische Ente* (“Turkish duck”) and, in French, as *canard de Barbarie* (i.e., from the Barbary coast of North Africa). Both these names obviously suggest that the New World origin of the bird was not well understood in Europe at the time. The eighteenth century seems to have marked the high point of popularity of the Muscovy, which, in many cases, was raised by the rural gentry mainly as a curiosity and as an ornament, just as they did with peacocks.

In France, where a restless revision of the culinary imagination has a glorious history, the bird became a meat animal, and it was the French who crossed the Muscovy with the common duck to produce the *mulard* duck. This vigorous duck hybrid has been force-fed grain to produce hypertrophied livers used in making *pâté de foie gras*. Especially in southwestern France, farms specializing in the production of this delicacy keep both geese and *mulard* ducks.

Elsewhere in the world where they were eventually introduced, Muscovies never acquired much importance in domestic economy. Possible exceptions to this generalization are parts of West Africa, because the tropical origins of Muscovy ducks made them better adapted to hot conditions than other poultry. In Asia, where duck husbandry is important in furnishing sources of animal protein, this relative from the New World has found no real acceptance. Muscovies

are an uncommon species in China, which has more ducks than any other country. Failure of the Muscovy to compete is tied in part to its lower productivity, in terms both of egg laying and body size. In addition, it does not fare as well in cold climates as the common duck. In fact, these are the reasons that commercial duck raisers in North America do not use this species. Placed on a world scale, *C. moschata domestica* has been among the less successful domesticated birds.

Daniel W. Gade

Bibliography

- Donkin, R. A. 1989. *The Muscovy duck, Cairina moschata domestica: Origins, dispersal and associated aspects of the geography of domestication*. Rotterdam/Brookfield, Vt.
- Gilmore, Raymond M. 1963. Fauna and ethnozoology of South America. In *Handbook of South American Indians*, ed. J. H. Steward, 345–464. New York.
- Nogueira-Neto, Paulo. 1973. *A criação de animais indígenas vertebrados*. São Paulo.
- Whitley, G. R. 1973. The Muscovy duck in Mexico. *Anthropological Journal of Canada* 11: 2–8.

II.G.18 🐦 Pigeons

Because humans seem to have had omnivorous ancestors, birds were probably a significant item in the human diet well before historic time. Many kinds of birds can be caught readily, and young adults are generally considered to be superior fare (Cott 1946). Additionally, the eggs of many kinds of birds are highly prized (Cott 1954). In the case of pigeons, the rock pigeon (*Columba livia*) is known to have frequented regions inhabited by our ancestors more than 300,000 years ago (Tchernov 1968).

At present, the domestic chicken (*Gallus domesticus*) and their eggs are extremely important in diets across the globe. Indeed, the husbandry of chickens is economically unequaled by that of any other bird. But the economic and dietary importance of domestic fowl is a relatively recent development and in large part reflects changes brought about by twentieth-century biology.

The domestic fowl, derived from Asian species of the genus *Gallus*, was introduced to the West and became known to people of the Mediterranean Near East only a little more than 2,500 years ago (Wood-Gush 1985). This was perhaps 3,000 years after rock pigeons were domesticated in southwestern Asia (Sossinka 1982) and long after they had become important in human diets. However, because pigeons also figured in early religions of the eastern Mediter-

anean region, the fact that they were domesticated birds does not explicitly address their use as food. Thus, the dual role pigeons have played in human affairs needs to be distinguished whenever possible.

The family Columbidae is widely distributed (Goodwin 1983). There are some 300 kinds of pigeons known to biologists, and many have been and still are used for food by humans worldwide. The record of dietary use is, however, without detail for most such species, as for example, doves. But it is likely that the flesh of most (although not necessarily their eggs) is suitable for human consumption (Cott 1946, 1954).

Rock Pigeons

The rock pigeon has a history in part coincident with that of humans for at least the past 12,000 years. The earliest information is of two kinds – the organic, sub-fossil, bony remains of pigeons in midden heaps in caves and the slightly later cultural record of human-pigeon interactions.

Bones from midden heaps, which constitute reasonably direct evidence that humans caught and ate rock pigeons, are present in eastern Mediterranean sites – caves in Israel that were used as dwellings by humans around 11,000 to 12,000 years ago. Because the bones are indistinguishable from those of wild rock pigeons currently living in the Near East (Bar-Yosef and Tchernov 1966), it is possible that immediate post-Pleistocene pigeons might have nested in caves used by humans. And it is also possible that their bones were incorporated into the midden heaps as a result of death by other than human agents. It is considerably more likely, however, that bones in these refuse piles indicate that the pigeons in question served as food for human hunters.

But so long as humans relied on catching wild pigeons – probably squabs from nests – pigeon would have been merely an occasional item of diet. Only after pigeons were domesticated and relegated to a life of confinement in cages did they make a regular appearance on the table. The first such attempts are not recorded, but rock pigeons are readily domesticated (Darwin 1868), and as the following discussion shows, this practice doubtless occurred relatively early.

The earliest evidence of domestication has been found in Sumerian statuary and cuneiform lists; in the remains at a funerary feast in a tomb at Saqqara, Egypt; in Sumerian culture, which includes a version of the Mesopotamian Flood Myth featuring a pigeon; and in small clay religious shrines as well as on sculptures depicting fertility goddesses, Astarte or Ishtar. Slightly later, there are Egyptian bird sacrifices on record and a Babylonian scribal list of bird names in cuneiform. We will briefly explore this evidence, beginning with that which dates from earliest times.

The Sumerian Flood Myth can be dated from a massive flood that occurred about 6,000 years ago (Langdon 1931). In the myth, a “dove” was the first bird

released from the still floating Ark, and it returned for the night, signifying no available dry land.

The early Egyptian material includes traces of pigeon bones from a funeral dinner apparently held some 6,000 years ago (Hansell and Hansell 1992). This is the first, and very dim, indication of what may have been domestic pigeons playing a role in human diets.

Sumerian clay shrines, dating from some 4,500 years ago, depict pigeons that were evidently of religious importance. There are also fertility figures from the same period that have pigeons perched on their uplifted arms (Langdon 1931). Quite possibly, the significance of the pigeons in this instance has to do with their unusual reproductive capabilities (the ability of captives to lay eggs and rear young in midwinter).

Around 4,000 years ago it was recorded that the Egyptian King Rameses II offered 58,810 pigeons to the god Ammon, at Thebes (Hansell and Hansell 1992). The large number of birds involved can only mean that Egyptians were practicing pigeon husbandry at that time, for there is no way in which wild rock pigeons could have been rounded up in such numbers.

An Old Babylonian cuneiform tablet dated from 3,830 to 3,813 years ago contains 70 Akkadian bird names in more or less taxonomic groups (Black and Al-Rawi 1987), and no fewer than 3 of the names refer to pigeons, one of which almost certainly applied to the rock pigeon.

A considerable summary of such ancient records treating pigeons as sacred or at least of religious importance may be found in the opening pages of W. Levi (1974). Unfortunately, little of the archaic information provides a positive identity for the early species of domesticated pigeon. Most likely, however, rock pigeons and at least one other species, probably a kind of turtledove, were involved. In addition, of the species of pigeons (and doves) still present in the Near East, the rock pigeon, wood pigeon (*Columba palumbus*), stock dove (*Columba oenas*), and turtledove (*Streptopelia turtur*) were surely used as food and may be among the birds represented on the clay shrines and in the lists just discussed.

Rock pigeons have been common, year-round residents in the Near East. They have extremely long breeding seasons, nesting semicolonially in caves and on ledges of coastal and montane cliffs, and are thought to have sometimes become synanthropic in early Near Eastern cities (Glutz and Bauer 1980). Their reproductive capabilities of persistent, serial breeding, beginning early and ending late in the year, when no other kinds of birds are active, would have been readily observed and, thus, part of the knowledge of the time. Such knowledge would have been of considerable importance in fostering efforts aimed at holding pigeons as captives.

Fragmentary evidence as well as outright speculation about human behavior suggests that the domestication of rock pigeons took place between 6,000 and 10,000 years ago. This must have occurred after the development of grain farming (around 10,000 years

ago), which was the major human cultural acquisition that permitted animal domestication in the Near East (Harlan and Zohary 1966). But although cereal agriculture and early animal domestication originated in the Near East, the wild rock pigeon has a large geographic distribution (Cramp 1985) overlapping early human agriculture in many regions of Europe and Asia (Piggott 1965; Ammermann and Cavalli-Sforza 1971).

Consequently, rock pigeons could easily have been domesticated at many different places and times in southeastern Europe, North Africa, the Near East, and southwestern and southern Asia. Doubtless, another cultural factor that assisted in the domestication of pigeons was the experience gained by humans in domesticating other creatures. Wolves, for example, were domesticated in the Near East at least 9,000 years ago and even earlier elsewhere (Reed 1969).

Pigeon chicks taken at about 2 weeks of age would have been somewhat less wild than their parents and more likely to have accepted captivity. Wild adult pigeons are nervous captives, do not accept cages, and thus sometimes injure themselves; they also spend less time in reproduction than birds accustomed to confinement. In light of this, one suspects that domestication could have occurred in one of two, somewhat different, ways.

The first would have been for hunters to have captured rock pigeon squabs in nesting caves, with the idea of holding them and feeding them in cages until they grew larger. Gradually, however, humans would have learned to continue such captivity until the birds matured and were able to reproduce. A second scenario would have humans erecting nesting platforms in dove caves. This would have turned them into primordial dovecotes, where the birds could have been harvested and where much could have been learned about their reproductive behavior and ecology.

Concurrent with these scenarios, early cities of the Near East may have been voluntarily colonized by wild rock pigeons. The mud and stone walls of such settlements would have made quite suitable nest sites, with nearby fields of wheat and barley providing nutrition. These synanthropic birds could subsequently have been taken into confinement by either of the two ways just mentioned.

Selection of Pigeons for Food

Captive pigeons probably were unconsciously selected by humans for tameness and reproductive vigor, but they would have been consciously selected for characteristics that increased their utility as high-quality food. The major one of these traits to be modified was size. Wild rock pigeons in the Near East are relatively small and average about 300 grams in undressed weight (Cramp 1985). But today, males of large domestic strains are known to reach more than seven times that size (Levi 1974). Thus, the mating of large individuals would readily have repaid any early pigeon keeper.

We have no documentation for such a selection process, but it is clear that it was practiced. By Roman

times, for example, a large-bodied domestic strain (usually identified with today's "runt") was already developed (Levi 1974). The runt is preserved in a variety of substrains by pigeon fanciers and commercial squab producers.

At the time that pigeons were developed for the table, smaller doves were probably also being selected. Dovecote pigeons are domestics that are, for the most part, allowed to fly freely in order to obtain their own food; they are also highly site-specific and return daily to the pigeon house or dovecote in which they were reared. Dovecotes are small buildings, usually raised above the ground and providing quarters for roosting and nesting pigeons as well as access for pigeon keepers to harvest the squabs.

Prior to the time that larger mammals were kept overwinter, pigeons were a reliable source of winter protein for rural people (Murton and Westwood 1966). Indeed, well-fed pigeons with a characteristically high fat content in muscle would have been important to such diets (Barton and Houston 1993). Pigeons have continued to be used for food up to the present time, although recent practice emphasizes the gourmet table rather than the farm kitchen (Levi 1974). As already mentioned, the current dominance of domestic fowl in human diets is a reflection not of its superior quality as a source of animal protein but of its greater adaptability to human poultry husbandry, which guarantees producers large returns on their work or investment.

Pigeon Husbandry in Europe and America

In earlier times, once a pigeon house was constructed, pigeons were easy to keep, at least at low to mid latitudes. The birds of a columbarium fed themselves on wild seeds and waste agricultural grains, and if their food was supplemented in winter, a large number – thousands of pairs – could be maintained throughout the year.

It is not clear at what time pigeon husbandry on such a large scale was found in Europe, but substantial columbariums were in operation in medieval times. Of the ones that are documented, some, such as those set up by thirteenth-century emperor Frederick II, were reasonably elaborate. The emperor's holdings covered southern Italy and eastern Sicily, which were dotted with more than a dozen of his castles.

Frederick moved from one to another of these castles in the course of a year, because the large size of his court regularly required a fresh supply of food. Frequent moves also provided time at any one castle for the satisfactory recycling of wastes in between visits. Each castle had a large dovecote, built within and sharing some of the castle walls, which the court relied on for a significant fraction of its fresh meat (Frederick II 1942). We know these details of the emperor's life because he was an ornithologist as well as a builder of castles. He left behind his writings on birds and the remains of his buildings, which clearly show that the columbarium was an integral part of the castles he designed.

While in Syria, where he was otherwise engaged on a Crusade, Frederick was able to secure novel domestic strains of pigeons. Birds from the Near East may not have been bred for table use, owing to religious practices of the Syrians, but the record is important to us because it suggests widespread pigeon husbandry at that time.

In post-Renaissance Europe, pigeon keeping was, to some extent, restricted to the privileged classes of society – to manor house lords and members of the clergy (Cooke 1920). Many large flocks were maintained, and recent estimates of numbers of dovecote pigeons in England and France in the sixteenth century run to the millions (Cooke 1920; Murton and Westwood 1966). As noted earlier, dovecote birds generally foraged for themselves in the agricultural countryside, going distances of perhaps 20 to 30 kilometers (Ragionieri, Mongini, and Baldaccini 1991).

As a consequence, peasant farmers on the estates and beyond sometimes had serious problems in getting their grains and pulses to sprout because of pigeon depredations. Maturing and ripe grain was also at risk, as was stored grain, for pigeons entered storage sheds if these were not secure. To compound the problem, farmers were prohibited from killing the birds. Indeed, some political historians have suggested that the victimization of farmers by the operation of aristocratic dovecotes contributed to the rebellion against social privilege culminating in the French Revolution of 1789 (Cooke 1920).

Whether true or not, such a sequestering of rights to rear pigeons by the privileged classes of medieval and post-Renaissance Europe speaks loudly of the significance of pigeons in the diet of people of preindustrial Europe and Asia. The birds were especially important in adding fresh protein to wintertime diets for those living at high latitudes or away from seacoasts, or anywhere that cattle could not regularly be overwintered. It is likely that chickens also provided such protein in the cold months (as well as at other times of the year), but chicken reproduction dropped off considerably in winter. Thus, wintertime diets that featured chicken depended on mature birds, in contrast to the use of pigeon squabs, which appeared regularly throughout the year.

Current Pigeon Husbandry

The current architecture of pigeon houses in the Middle East, Mediterranean Europe, and North Africa may be little changed from earlier times, and rural Egyptians still build tall, earthen dovecotes – towers of mud into which clay pots have been placed for pigeons to nest in (Hollander 1959; Hafner 1993). By contrast, rural or small-town Italians use metal screening and planks of wood, usually keeping the birds in a flypen arrangement. But however pigeons are kept, adult birds of less than 7 years of age are employed as breeders, and their squabs are taken for the table at about 30 days of age (Levi 1974). The adults tend to overlap their broods, so that when well fed and fully

confined, they may produce 12 to 18 squabs per year (Burley 1980; Johnson and Johnston 1989). Thus, even a small colony of 10 pairs could provide a family with squabs for the table each week.

Dovecote pigeonry in the Americas was introduced by settlers from England, France, and Germany to early seventeenth-century Nova Scotia and Virginia (Schorger 1952). Pigeons for the table (as distinct from racing or show purposes) are kept in much the same fashion today as in earlier times. The major difference is that the pigeon house in America is generally a single-story building that is spread out over a very large area relative to those of historic Europe and Asia, which tended to be columnar (Cooke 1920; Hansell and Hansell 1992).

Because a dovecote or columbarium consists of a large number of boxes, cells, or breeding enclosures in which a pair of birds can build a nest and rear young, a western European columnar columbarium facilitates the work of cleaning up by concentrating the birds' droppings. In contrast, a ranch-style unit of pigeon houses, such as in North America, allows increased time intervals in such housekeeping. The houses are usually of wooden frame construction on a concrete pad; a center aisle allows entry by keepers to the nest boxes for feeding and housekeeping, as well as for keeping track of nests and eggs, marking squabs, and, ultimately, collecting them.

The birds may be confined, which means they live partly in a roofed nest-box enclosure and partly in a screened flypen, open to the elements. The classical dovecote operation was employed until relatively recently, even in commercial establishments (Levi 1974), but it is now found chiefly on single-family farms.

Use as Food

Squabs prepared for the table are about as large as small chickens, but their distribution of edible flesh is different. In contrast to chickens, pigeons are strong-flying birds, and perhaps 30 percent of a squab's overall weight, and 70 percent of the muscular weight, will be in the paired pectorales or flight muscles. Pigeons, although also adept at running and walking, have leg muscles that are proportionally smaller than those of chickens.

Pigeons appear in the world's first cookbook, which is attributed to Apicius, a first-century Roman, but is thought to have originally been a collection of Greek monographs on cookery by several authors (Vehling 1936). If we judge only by the frequency of its pigeon recipes, which number 2 against 18 for chicken, then it would seem that the ascendancy of the latter in human diets was already marked in Mediterranean Europe by the first century. However, since all the chicken recipes could be as readily done with pigeon, it may be that the authors were offering us their own preferences (an author's prerogative) rather than indicating the relative availability of the

two sorts of birds. The recipes for pigeon are ones dominated by raisins, honey, and dates, suggesting that perhaps the more pronounced flavor of pigeon is better able than chicken to emerge from such sugary dishes.

In any event, today pigeons are regularly prepared for the table in the same manner as domestic fowls, but because of the more assertive flavor of pigeons, stronger spicing may be used if desired. Pigeon breast muscles are "dark meat," so that a significant fraction of current recipes employ plenty of garlic, and those using wines specify full-bodied reds, or sherries and other fortified wines.

Pigeons Other Than Rock Pigeons

Domesticated rock pigeons are the most important food species of the 300 kinds of pigeons in the world. Only one other, the ringed turtledove (*Streptopelia risoria*), has been domesticated, but it is smaller than the rock pigeon and not of general significance for human diets. A large number of other kinds of pigeons of variable sizes are also used on occasion for food by humans. Many of these are secured by hunters using guns or nets, such as the North American mourning dove (*Zenaidura macroura*), the European wood pigeon (*C. palumbus*), and many other Asiatic, African, South American, and Australian species.

Such hunting may be regulated by government restrictions, so that a yearly harvest occurs each autumn when the birds are in fairly dense migratory flocks. But hunting may also be restricted to autumn for economic reasons, as this is frequently the only time in which it is profitable to secure the birds. A notable example is the famous netting operations in the valleys of the Pyrenees between France and Spain – valleys used by a large fraction of European wood pigeons in migration.

Unfortunately, the passenger pigeons (*Ectopistes migratorius*) of North America were hunted at any time of year and became extinct around 1900, although it should be noted that the cutting of the North American hardwood forest, which destroyed a significant food source and was the chief habitat of the birds, was an equally important factor (Bucher 1992). Nonetheless, it is fair to say that pigeons and their relatives comprise a group inordinately disposed toward extinction at the hands of human hunters. No fewer than 15 species have been exterminated in historic time, including such remarkable birds as the dodo (*Raphus cucullatus*) and the solitaire (*Pezophaps solitaria*), as well as the passenger pigeon. In all instances it is clear that the birds were used for food by humans, but it is a fact that habitat destruction and the introduction of cats, dogs, and rats also contributed to such extinctions, most of which occurred on islands.

Richard F. Johnston

Bibliography

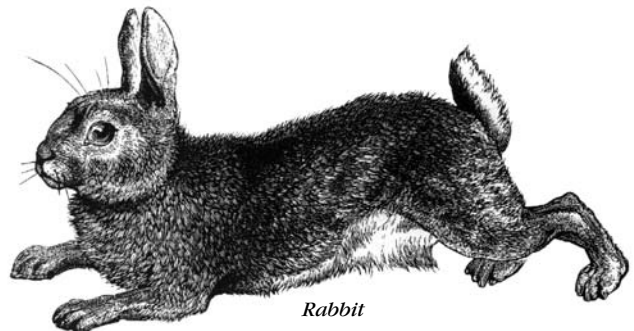
- Ammermann, A., and L. Cavalli-Sforza. 1971. Measuring the rate of spread of early farming in Europe. *Man* 6: 674-88.
- Barton, N. W. H., and D. C. Houston. 1993. A comparison of digestive efficiency in birds of prey. *Ibis* 135: 363-71.
- Bar-Yosef, O., and E. Tchernov. 1966. Archaeological finds and the fossil faunas of the Natufian and Microlithic industries of the Hyonim Cave (Western Galilee, Israel). *Israel Journal of Zoology* 15: 104-40.
- Black, J. A., and F. N. H. Al-Rawi. 1987. A contribution to the study of Akkadian bird names. *Zeitschrift für Assyriologie und Vorderasiatische Archäologie* 77: 117-26.
- Bucher, E. H. 1992. The causes of extinction of the passenger pigeon. *Current Ornithology* 9: 1-36.
- Burley, N. 1980. Clutch overlap and clutch size: Alternative and complementary tactics. *American Naturalist* 115: 223-46.
- Cooke, A. 1920. *A book of dovescotes*. London.
- Cott, H. B. 1946. The edibility of birds. *Proceedings of the Zoological Society of London* 116: 371-524.
1954. The palatability of the eggs of birds. *Proceedings of the Zoological Society of London* 124: 335-463.
- Cramp, S. 1985, ed. *Handbook of the birds of Europe, the Middle East and North Africa: The birds of the Western Palearc*. Vol. 4. Oxford.
- Darwin, C. 1868. *The variation of animals and plants under domestication*. London.
- Frederick II. [1250] 1942. *The art of falconry*, trans. C. A. Wood and F. M. Fyfe. Stanford.
- Glutz von Blotzheim, U. N., and K. M. Bauer. 1980. *Handbuch der Vögel Mitteleuropas*, Vol. 9. Wiesbaden, Germany.
- Goodwin, D. 1983. *Pigeons and doves of the world*. Ithaca, N.Y.
- Hafner, D. 1993. *A taste of Africa*. Berkeley, Calif.
- Hansell, P., and J. Hansell. 1992. *A dovescote heritage*. Bath, England.
- Harlan, J., and D. Zohary. 1966. The distribution of wild wheats and barleys. *Science* 153: 1074-80.
- Hollander, W. F. 1959. Pigeons in ancient Egypt. *American Pigeon Journal*. August: 254-5.
- Johnson, S., and R. Johnston. 1989. A multifactorial study of variation in interclutch interval and annual reproductive success in the feral pigeon, *Columba livia*. *Oecologia* 80: 87-92.
- Langdon, S. 1931. *The mythology of all races. V. Semitic*. Boston.
- Levi, W. 1974. *The pigeon*. Sumter, S.C.
- Murton, R., and N. Westwood. 1966. The foods of the rock dove and feral pigeon. *Bird Study* 13: 130-46.
- Piggott, S. 1965. *Ancient Europe from the beginnings of agriculture to classical antiquity*. Chicago.
- Ragionieri, L., E. Mongini, and E. Baldaccini. 1991. Problemi di conservazione in una popolazione di colombo sylvatico (*Columba livia livia* Gmelin) della Sardegna. In *Atti del convegno genetica e conservazione della fauna*, ed. E. Randi and M. Spagnesi, 35-46. *Supplemento alle Ricerche Biologia della Selvaggina* 18: 1-248.
- Reed, C. A. 1969. Patterns of animal domestication in the prehistoric Near East. In *The domestication and exploitation of plants and animals*, ed. P. J. Ucko and G. W. Dimbleby, 361-80. London.
- Schorger, A. W. 1952. Introduction of the domestic pigeon. *Auk* 69: 462-3.
- Sossinka, R. 1982. Domestication in birds. *Avian Biology* 6: 373-403.
- Tchernov, E. 1968. Preliminary investigation of the birds in the Pleistocene deposits of 'Ubeidiya. *Publications of the Israel Academy of Sciences and Humanities*, 1-45.
- Vehling, J. D. 1936. *Apicius cookery and dining in imperial Rome*. Chicago.
- Wood-Gush, D. G. M. 1985. Domestication. In *A dictionary of birds*, ed. B. Campbell and E. Lack, 152-4. Vermillion, S. Dak.

II.G.19 🐰 Rabbits

Rabbit production is significant in several countries, especially France, Italy, Malta, and Spain, where there is a long tradition of consuming rabbit meat. In the past, great numbers of rabbits were raised by subsistence farmers, who fed them locally collected forages such as weeds, grasses, and vegetable by-products. But with the intensification of agriculture, particularly in the twentieth century, and the decline in "peasant farmers," rabbit production as a cottage industry declined. However, because the tradition of eating rabbit meat endured in western European countries, an intensive, commercial, industrial-scale production of rabbits has developed to meet continuing demand.

Both the origins and the evolution of the domestic rabbit are difficult to trace. Rabbits are in the order Lagomorpha, which dates back about 45 million years in the fossil record to the late Eocene period. Modern lagomorphs (rabbits and hares) belong to two families (Leporidae and Ochotonidae) consisting of 12 genera. They range from the highly successful hares and rabbits of the *Lepus*, *Oryctolagus*, and *Sylvilagus* genera to several endangered genera and species. Although rabbits and hares appear to have originated in Asia, all breeds of domestic rabbits are descendants of the European wild rabbit (*Oryctolagus cuniculus*) and are Mediterranean in origin. There are more than 100 breeds of domestic rabbits, ranging in size from dwarf breeds with an adult weight of less than 1 kilogram (kg) to giant breeds weighing in excess of 10 kg.

The first recorded rabbit husbandry has been dated to early Roman times, when rabbits were kept in *leporia*, or walled rabbit gardens. They reproduced in these enclosures and were periodically captured and butchered. During the Middle Ages, rabbits were similarly kept in rock enclosures in Britain and western Europe. True domestication, which is believed to have



Rabbit

taken place in the Iberian Peninsula, probably began in the sixteenth century in monasteries. By 1700, several distinct colors had been selected. There are now many different coat colors (for example, agouti, tan, brown, white, blue, black, and red) and coat types (such as angora, normal, rex, satin, and waved), which provide for great diversity in the color and texture of the fur. (The coat-color genetics of rabbits have been reviewed comprehensively by McNitt et al. 1996.)

Domestic rabbits are now raised in virtually all countries, although, in some notable instances, feral rabbits have become major pests. Beginning in the Middle Ages, sailors introduced these animals into islands along various sea lanes – to be used as a source of food – and wherever rabbits were released, they increased greatly in number at the expense of indigenous plants and animals. In 1859, a single pair of European wild rabbits was taken to Australia, and within 30 years, these had given rise to an estimated 20 million rabbits. Feral rabbits also became a serious problem in New Zealand, which (like Australia) offered a favorable environment, abundant feed, and an absence of predators.

Periods of peak interest in rabbit production have coincided with times of economic hardship or food scarcity, both of which have encouraged people to produce some of their own food. In the United States and Europe, such periods during the twentieth century have included the Great Depression of the 1930s and World War II. A few rabbits kept in the backyard and fed weeds, grass, and other vegetation can provide a family with much inexpensive meat, and keeping them is practically trouble-free. Not only do they eat homegrown feeds, rabbits do not make noise and are easily housed in small hutches. When economic times improve, however, the interest in home production of food wanes, and those who ate rabbit meat under conditions of deprivation tend during better times to regard it as a “poor people’s food.”

Rabbits have a number of biological advantages when raised for meat production (Cheeke 1986, 1987). They experience rapid growth, reaching market weight at 8 to 10 weeks following birth. Their rate of reproduction is high: Theoretically, with immediate postpartum breeding, females (called does) can produce as many as 11 litters per year. And, as already noted, rabbits can be raised on fibrous feedstuffs, forages, and grain-milling by-products (such as wheat bran) and thus do not require high-quality feed grains.

Rabbit meat is a wholesome, tasty product. Compared to other common meats, it is high in protein and low in fat calories, cholesterol, and sodium (Table II.G.19.1) – properties that are related to the animals’ low-energy, high-fiber diets. Because they are not fed much in the way of grain, rabbits do not have excess energy (calories) to store as body fat.

Total world production of rabbit meat is estimated to be about 1 million tonnes per annum (Lukefahr and Cheeke 1991). On a per capita basis, the major

Table II.G.19.1. *Nutrient composition of rabbit meat*

Nutrient	Amount of nutrient
Crude protein (%)	18.5 ^a
Fat (%)	7.4 ^a
Water (%)	71 ^a
Ash (%)	0.64 ^a
Unsaturated fatty acids (% of total fatty acids)	63
Cholesterol (mg/100 g)	136 ^b
Minerals^b	
Zinc (mg/kg)	54
Sodium (mg/kg)	393
Potassium (g/kg)	2
Calcium (mg/kg)	130
Magnesium (mg/kg)	145
Iron (mg/kg)	29
Vitamins^b	
Thiamine (mg/100 g)	0.11
Riboflavin (mg/100 g)	0.37
Niacin (mg/kg)	21.2
Pyridoxine (mg/kg)	0.27
Pantothenic acid (mg/kg)	0.10
Vitamin B ₁₂ (µg/kg)	14.9
Folic acid (µg/kg)	40.6
Biotin (µg/kg)	2.8
Amino acids^c	
Leucine	8.6
Lysine	8.7
Histidine	2.4
Arginine	4.8
Threonine	5.1
Valine	4.6
Methionine	2.6
Isoleucine	4.0
Phenylalanine	3.2

^aWet weight basis.

^bDry weight basis.

^cAmino acids expressed as percentage of protein.

Source: Adapted from McNitt et al. (1996).

rabbit-consuming nations are those already mentioned as leading producing countries – Malta, France, Italy, and Spain – where rabbit has traditionally been an important meat. But in spite of many attempts to develop rabbit production and consumption in other areas, such efforts have largely been unsuccessful, which is especially unfortunate in developing countries, where consumption of good-quality protein is generally low. Moreover, the diet that rabbits consume in no way places them in competition with humans for food. They can be raised in simple structures, and the carcass size is small, so that meat storage with refrigeration is unnecessary.

Nonetheless, despite numerous rabbit development programs in Africa, Asia, and Latin America, there are few if any examples of permanent success following their introduction into local farming systems, with

a number of factors accounting for this lack of success. Rabbits are quite susceptible to heat stress, so their performance in tropical countries is poor. Their Mediterranean origin has best suited them for hot, arid climates (for example, northern Africa) rather than hot, humid areas (such as equatorial Africa). In addition, because of the tremendous problems with rabbits following their introduction into Australia, many countries are reluctant to promote rabbit production out of fear that they will encounter similar problems. Concern is also often expressed about zoonoses – diseases that might be transmitted from animals to humans. In the case of rabbits, this fear is especially acute because of the disease tularemia, which humans can acquire when handling wild rabbits. In fact, however, tularemia is not a significant hazard with domestic rabbits. Finally, there is the usual cultural resistance to a new source of meat. Although there are no religious taboos against eating rabbit meat, there is the common perception that the animals are “cute” or “warm and fuzzy,” which initiates an aversion in many people to rabbit consumption.

Another problem with disease has to do with myxomatosis. In Britain as well as in Australia, the hunting of wild rabbits and their sale in butcher shops was formerly widespread. But then, in both countries, the need to control the animals led to the introduction of myxomatosis, which is a devastating viral disease causing grotesque facial lesions and swelling, oral bleeding, and a generally distressing appearance. The sight of large numbers of wild rabbits dying with these symptoms hardly whets the appetite for rabbit meat, and needless to say, the traditional marketing and consumption of this meat in countries where myxomatosis was introduced has fallen drastically.

The principal U.S. market for rabbit meat is gourmet restaurants, where it is served as a traditional French specialty. In addition, ethnic markets patronized by European immigrants of French and Italian origin also are significant in large metropolitan areas, especially on the East and West coasts.

By way of conclusion, rabbit production is significant in several Mediterranean countries, where the domestic rabbit originated and was particularly well suited for small-scale production on subsistence or peasant farms. Even in these countries, however, with the introduction of fast-food restaurants and more American-style eating, there is a definite trend of decreasing rabbit consumption, especially among young people, suggesting that the importance of the rabbit as a meat animal will continue to decline. In the United States, strong interest in animal rights and vegetarianism among young people probably means that backyard rabbit production, with home slaughter of the animals, is unlikely to have much appeal. A small rabbit industry, with the production of meat for specialty or gourmet restaurants, is likely the main future of rabbit production in most countries.

Peter R. Cheeke

Bibliography

- Cheeke, P. R. 1986. Potentials of rabbit production in tropical and subtropical agricultural systems. *Journal of Animal Science* 63: 1581–6.
1987. *Rabbit feeding and nutrition*. San Diego, Calif.
- Lukefahr, S. D., and P. R. Cheeke. 1991. Rabbit project development strategies in subsistence farming systems. *World Review of Animal Production* 68: 60–70.
- McNitt, J. I., N. M. Patton, S. D. Lukefahr, and P. R. Cheeke. 1996. *Rabbit production*. Danville, Ill.

II.G.20 🐢 Sea Turtles and Their Eggs

From earliest times the seashore, with its rich and diverse marine fauna, has been a uniquely attractive environment for humans, providing them with accessible, palatable, and protein-rich sustenance (Sauer 1962). Among other foods – at least in warmer seas of tropical latitudes – were giant marine turtles. These could be harvested with relative ease, either on the beaches while nesting or netted or harpooned offshore. Their soft-shell and Ping-Pong-Ball-like eggs, deposited in clutches of 100 or more in the warm sand of favored beaches, provided further nutritious fare.

Among the six or seven species of giant marine reptiles that are recognized, it is the green turtle, the *Chelonia mydas* of turtle-soup fame, that has contributed most to the human diet. Unlike other sea turtles, the greens are exclusively herbivores, thus accounting for the savory quality of their veal-like flesh. Their name derives not from the color of their shell or skin but from the soft, greenish gelatinous material known as calipee, found beneath the plastron (lower shell) and scraped from slaughtered turtles to prepare the thick green soup renowned among gastronomes.

The flesh of other sea turtles, although eaten by some coastal peoples, is generally reputed to have a somewhat uninviting fishy taste. If the animal has ingested toxic algae or crustaceans, it may even be poisonous. The hawksbill (*Eretmochelys imbricata*), prized from antiquity as the source of the beautifully mottled tortoiseshell of commerce, is valued for jewelry and ornamentation, whereas the olive ridley (*Lepidochelys olivacea*) has recently been much sought after for its skin, used to make leather goods. This use has been stimulated by changing styles and a scarcity of crocodile skins. The smaller Kemp's ridley (*Lepidochelys kempii*), the loggerhead (*Caretta caretta*), and the giant leatherback (*Dermochelys coriacea*) have not traditionally been utilized. The eggs of all species are eaten. Egg collection is either an open or a clandestine business on numerous strands throughout the tropical world.

The countless "turtle islands" (such as Islas Tortugas, Iles Tortues, and Schilpad Eilanden) of tropical seas bear witness to the preference of green and other sea turtles for uninhabited offshore islands for nesting as well as to their remarkable fidelity to specific, and sometimes quite small, beaches. They unerringly return to these sites after migrations to feeding areas that may be a thousand or more miles away. Such nesting habits, especially pronounced among the greens and the two ridleys, have made sea turtles particularly vulnerable to human exploitation and lie at the base of their endangered status.

The males spend their entire lives at sea, but the females, once gravid, may make several nocturnal visits to their ancestral beaches during each nesting cycle. These cycles usually come at intervals of four or more years. Scooping out a pit in the sand with their flippers, they bury their one hundred or so eggs before lumbering back to the surf and waiting mates. Two months later, the frenzied hatchlings, not much larger than a silver dollar, emerge to scamper quickly toward the water through a gauntlet of waiting predators. For those that survive, this is the beginning of a long and mysterious migration. As mature adults they will return several years later to the same beach on which they were born. Their imprinting, and the guidance mechanism that allows them, without visible landmarks, to travel the great distances between feeding areas and nesting beaches, remains among nature's grandest enigmas.

Atlantic and Pacific populations, long separated, carry distinctive DNA markers. Moreover, size helps to distinguish the Atlantic turtles from their counterparts on the West Coast of America. The latter constitute the subspecies *Chelonia mydas agassiz*, known locally as *caguama prieta* (black turtle), whose adult members weigh from 65 to 125 kilograms. By contrast, their Atlantic and Indian Ocean counterparts are substantially larger (100 to 200 kilograms), with some from Ascension Island in the past weighing a reported 600 kilograms.

Today, sea turtles are disastrously overexploited by a rapidly expanding world population and are victimized by pollution, coastal development, and new fishing technologies. Worldwide concern for their future has led to the recent imposition of tight controls on sea turtle exploitation by all but traditional fishing folk. In the last few years, all marine turtles save the Australian flatback (*Natator depressus*) have been classified as either "threatened" or "endangered," and traffic in them and products made from them have been largely eliminated through the Convention on International Trade in Endangered Species (CITES), subscribed to by most of the world's nations. With the newly sharpened sensitivity toward conservation, and with trade in turtle products curbed or banned, the days of turtle soup and turtle steaks in the gourmet restaurants of the world's great cities have effectively come to an end.

Old World Cultural Attitudes

Cultural attitudes toward foods, particularly the consumption of animal flesh, may be decisively influenced by religion. Avoidance of sea turtle meat, as with that of the freshwater or land tortoises with which they are often confused, is widespread in South and Southeast Asia (Simoons 1994; Charles Tambiah, personal communication). The eggs are at the same time, much sought after, prized for their presumed health-giving and aphrodisiac properties. Hinduism holds turtles in veneration. Lord Vishnu is said to have taken the form of a sea turtle during one of his reincarnations, raising the world from chaos and conflict. Turtles are thus depicted as bearers of the world and, as such, command respect. Among the devout the meat is not consumed.

In theory, the Islamic faith prohibits eating the meat of reptiles. However, although this is a restriction affecting vast numbers of shore people around the Indian Ocean, it is apparently not operative in North Africa. But, although the Muslim Malay may not eat turtle, the many Chinese living in Singapore and elsewhere in the area have no aversion to it. Turtles offered in urban markets may be taken in nets or harpooned, but turning them over on the beaches is prohibited.

Buddhists, too, avoid turtle flesh, and to gain favor with the deity, they set free turtles that become entangled in their fishing nets. The Burmese are said to consider turtles divine and keep them in tanks on pagoda grounds where they are fed special foods, but this practice may more often involve river turtles. Early Chinese sources refer to freshwater or land tortoises as symbolic of the good and the long life. Only with the conquest of the south in the Han period (206 B.C. to A.D. 220) did sea turtles become generally available. In T'ang times (A.D. 618-907) the green sea turtle and its calipee are recorded as having been a tribute to the royal court paid by the city of Canton (Simoons 1991). In contemporary China, turtles apparently continue to occupy a special niche in folk belief and the apothecary trade. The recent world-record-shattering performances of several female Chinese track-and-field athletes have been attributed to their drinking of turtle blood (*Sports Illustrated*, October 24, 1994).

If there are many cases of abstinence, there are exceptions that test the rule. On Hindu Bali, as among Polynesian and Micronesian groups and converted Christians generally, turtle flesh is especially consumed at festivals and on ceremonial occasions. Pliny long ago wrote of a cave-dwelling people at the entrance to the Red Sea who worshiped the turtle as sacred yet ate its flesh. Among subsistence shore-dwelling communities of the Indian subcontinent and those throughout Southeast Asia on whom religions often rest easily, turtle is still likely to provide a significant and palatable dietary supplement as well as a source of occasional cash income. Pagan coastal peoples in Southeast Asia have generally held sea turtle eggs and flesh in high esteem. Among Australian Abo-

rigines who live close to the sea, the green turtle remains a principal totem.

At the same time, many, and perhaps most, of the turtle beaches of these southern seas support intensive egg-collecting operations under licensing systems controlled by local authorities. These systems are often designed to assure that sufficient quantities of eggs are left to support reproduction of the turtle populations. There were, for example, more than 30 such licensed areas for egg collecting not long ago on the east coast of Malaya. The three Turtle Islands off Sarawak until 1950 consistently yielded harvests of from 1 to 2 million eggs a year, the product of a population of perhaps 10,000 females (Hendrickson 1958). Watchers on each island marked new nests each night with flags, returning in the morning to dig and box the eggs for shipment to Kuching, the Sarawak capital. The proceeds went to charities or the mosques. The killing of sea turtles is prohibited in Sarawak, as in most South-east Asian countries, but poaching is widespread.

The American Experience

Among Native Americans encountered by the first Europeans in the turtle-rich Caribbean, attention appears to have been focused on the giant reptiles as a source of meat, whereas their eggs were of secondary interest. The green turtle was and still is at the base of the diet of such coastal people as the Miskito of Nicaragua and Honduras, the Baja California tribes, and the Seri of Sonora, all living close to major turtle pasturing grounds. In the West Indies, unfortunately, the large populations of nesting and grazing turtles described by the early chroniclers at Grand Cayman, the Dry Tortugas, and Bermuda were quickly exterminated (Carr 1954). Only at Tortuguero in Costa Rica and on tiny Aves Island off Venezuela do the greens continue to congregate in numbers in the Caribbean.

At least one Carib group in the Lesser Antilles was said not to have eaten sea turtle "for being fearful of taking on the characteristics of that reptile" (Rocheport 1606, 2: 202). Yet eggs were relished. A similar preoccupation with turtle eggs, rather than turtle flesh, was evident for early Indian peoples on the west coast of Mexico and in Brazil. Turtle eggs are smaller than those of poultry but have more fatty yolk. They are often consumed raw. One might speculate, in terms of conservation, whether it is better to take the turtles or their eggs. Had both been subject to unrestrained exploitation, the prospects for the survival of the species would have been bleak much earlier.

The Europeans who first came into contact with the green and hawksbill populations of the Caribbean were not of one accord in their judgment of this fortuitously accessible new resource. The Spanish and Portuguese seemed for the most part uninterested in turtle. Alvise da Cadamosto, the first Portuguese to mention what must have been the green turtle, fed it to his crew in the Cape Verde Islands in 1456 and found it "a good and healthy" food. Fernandez de

Oviedo y Valdez, in his *Historia Natural* of 1526, agreed. But most Spanish and Portuguese chroniclers of the early period ignored the animal or suspected it of being poisonous; it was the later-arriving English who were most outspoken in their praise of the green turtle's virtues (Parsons 1962). Its health-giving qualities were much commented upon by observers of the seventeenth and eighteenth centuries. To John Fryer (1909: 306), writing of East India and Persia in the late seventeenth century, it was "neither fish nor fowl nor good red herring, restoring vigor to the body and giving it a grace and luster as elegant as viper wine does to consumptive persons and worn out prostitutes."

Many an ill-disposed Englishman on Jamaica went to the Cayman Islands during the turtling season to recover his health by feasting on turtle. As a cure for scurvy and relief from the monotony of a hardtack and salt-beef diet, the meat was much prized by explorers, merchantmen, and buccaneers. The great clumsy creatures were abundant, easy to catch, and most important in the tropical heat, able to be kept alive on the decks of ships for weeks. The late Archie Carr (1973) suggested that the green turtle more than any other dietary factor supported the opening up of the Caribbean. It seems to have played a similar role in the Indian Ocean. William Dampier, that rough seaman who, Oliver Goldsmith observed, added more to natural history than half the philosophers who went before him, made repeated and extensive references to sea turtles as a shipboard meat reserve in his *Voyages*, written between 1681 and 1688 (Dampier 1906). In his eyes, the eggs were for natives.

Learning from the Miskito

The coastal Miskito are the world's foremost sea turtle people (Carr 1973; Nietschmann 1973, 1979). The coral cays and shelf off their Nicaraguan home coast are the principal feeding ground for green turtles from the renowned Tortuguero rookery, some 200 miles to the south in Costa Rica. Under subsistence exploitation regulated by local Miskito communities with strong cultural, religious, and economic ties to the species, the population remained stable. But with commercialization, first by Cayman Islanders who had seen the turtles of their own island decimated, and then by other foreign interests, extraction rates became excessive. Yet this coast, between Cape Gracias a Dios and Bluefields, still supports the largest remaining population of greens, and also hawksbills, in the Caribbean. With the recent establishment of the Miskito Coast Protected Area, there is prospect for a return to culturally regulated exploitation after a long period of overuse (Nietschmann 1991).

The Miskito may have taught the English to appreciate turtle. As early as 1633, a trading station had been established among the Miskitos at Cape Gracias a Dios by English adventurers from the Puritan colony at Old Providence Island. From the beginning, relations between natives and traders were amicable,

encouraging a sort of symbiotic relationship that was nurtured in part by mutual antagonism toward the Spaniard. The Indians, superb boatmen, had an "eye" for turtles that never ceased to amaze the Europeans. Many an English and Dutch pirate vessel carried at least one Miskito man as a "striker" to harpoon turtle for the mess table. "Their chief employment in their own country," wrote Dampier,

is to strike fish, turtle, and manatee. . . . for this they are esteemed by all privateers, for one or two of them in a ship will maintain 100 men, so that when we careen our ships we choose commonly places where there is plenty of turtle or manatee for these Miskito men to strike; it is very rare to find privateers without one or more of them. (Dampier 1906, 1: 39)

Caymanian turtlers were working the Miskito shore by at least 1837. From Grand Cayman, turtle boats could reach the cays in three or four days. The turtlers assembled their catch at temporary camps in the cays, carrying them north at the end of the season to be kept in "crawls" until marketed. From 2,000 to 4,000 turtles were taken annually. (A turtle-soup cannery was established in Grand Cayman in 1952 by the Colonial Development Corporation, but it closed after one year.) When, in 1967, this traditional arrangement with the Caymanian turtlers was terminated by the Somoza government and turtling rights were granted to higher bidders, the extraction rate soared to an insupportable 10,000 a year. By 1979, international conservation pressure had forced the Nicaraguan government to shut down the turtle companies and to ban further commercial exploitation (Nietschmann 1993).

Turtle was in as great demand as a slave food in the West Indian colonies in the seventeenth and eighteenth centuries as was salt cod from Newfoundland. But the reptile was also enjoyed by the West Indian white aristocracy. It was considered a special delicacy when eaten fresh. "To eat this animal is the highest perfection," wrote Goldsmith (1825: 164), "instead of bringing the turtle to the epicure, he ought to be transported to the turtle." Janet Schaw, describing her visit to Antigua in the 1770s, wrote:

I have now seen turtle almost every day, and though I never could eat it at home, am vastly fond of it here, where it is indeed a very different thing. You get nothing but old ones there [London], the "chickens" being unable to stand the voyage; and even these are starved, or at best fed on coarse and improper food. Here they are young, tender, fresh from the water, where they feed as delicately and are as great epicures as those who feed on them. (Schaw 1939: 95)

The special quality of turtle soup was said to be that it did not "cloy." In other words, one could eat almost any quantity without ill effects. Its easily assimilated proteins, without carbohydrate or fat, were pro-

claimed to prepare the stomach in superb fashion for what was to come. When banquets started with this soup, the diner was considered best able to enjoy the numerous rich dishes to follow. Goldsmith wrote that turtle "has become a favorite food of those who are desirous of eating a great deal without surfeiting. . . . by the importation of it alone among us, gluttony is freed from one of its greatest restraints" (1825: 674). The soup, flavored with sherry, capsicums, ginger, cloves, and nutmeg, and served piping hot, was considered at its fiery best "when, after having eaten, one is obliged to rest with his mouth wide open, and cool the fevered palate with Madeira or Port" (Simmonds 1883: 366). In 20 years in the West Indies, one doctor professed, he had never heard of an "accident" arising from eating it! It was also held to be an ideal food for convalescents, especially when served in jellied form.

The Dutch, although they partook of it, seem to have been rather indifferent to turtle in the East, perhaps because of their close association with the Malays, who avoided the meat. In the West, the French, while interested, found but a limited supply of green turtle available to them, most of the best turtling grounds being under English control. From the seventeenth-century account of Père Labat, a Dominican monk, it is evident that the animal's merits were not unrecognized. Yet it did not rate so much as a mention in Brillat-Savarin's exhaustive *Physiologie du gout*, written in 1825. For the French, turtle was clearly an English dish.

Spanish Disdain for Turtle

As rivalry between Spain and England intensified, Spanish disdain for what the English considered among the finest of foods heightened. Dampier (1906, 2: 399), describing the turtles found on the Brazilian coast, wrote in 1699:

neither the Spaniards nor Portuguese lov(e) them; Nay they have a great antipathy against them, and would rather eat a porpoise, tho' our English count the green turtle very extraordinary food. The reason that is commonly given in the West Indies for the Spaniards not caring to eat them is the fear they have lest, being usually foul-bodied, and many of them pox'd (lying as they do so promiscuously with their Negresses and other She-slaves), they should break out loathsomely like lepers; which this sort of food, 'tis said, does much incline men to do, searching the body and driving out any such gross humours.

Richard Walter, writing in 1748 while with Lord George Anson on his voyage around the world, thought it strange, considering the scarcity of provisions on the Pacific coast of Central America,

that a species of food so very palatable and salubrious as turtle should be proscribed by the Spaniards as unwholesome and little less than

poisonous. Perhaps the strange appearance of this animal may have been the foundation of this ridiculous and superstitious aversion, which is strongly rooted in all of the inhabitants of this coast. (Walter 1928: 208)

Of the Indians and Negroes (slaves of the Spaniards) who had been taken as prizes in Peru, Walter noted:

These poor people, being possessed with the prejudices of the country they came from, were astonished at our feeding on turtle and seemed fully persuaded that it would soon destroy us. . . . it was with great reluctance and very sparingly that they first began to eat it; but the relish improving upon them by degrees, they at last grew extremely fond of it, and preferred it to every other kind of food. . . . a food more luxurious to the palate than any their haughty Lords and Masters could indulge in. (Walter 1928: 288)

The Spaniards' apparent lack of interest in turtle appears in part a reaction to the close identification of it with the rival and hated English. In his study of Old World food prejudices, Frederick Simoons (1994) has shown the frequency with which particular animals or foods have become identified with particular ethnic, religious, or other groups through the course of history. The tendency to identify peoples with distinctive food habits is only a step from the rejection of foods simply because they are associated with a rival group. Pastoralists' rejection of the pig, an animal closely associated with and symbolic of the settled farmer, is an extreme, but by no means isolated, example of this sort of attitude.

The London Turtle Trade

Although the virtues of turtle had long been familiar to West Indian planters and to men of the sea, its introduction to the tables of London came only in the mid-eighteenth century. The *Gentleman's Magazine* in 1753 and 1754 carried several notices of large sea turtles, brought from Ascension Island and the West Indies, being dressed at public houses in London. One of the turtles was brought by Lord Anson. At the Kings Arms tavern in Pall Mall, the door of the oven had to be taken down to admit the plastron of a 350-pound specimen. "It may be noted," it was observed, "that what is common in the West Indies is a luxury here" (Anon. 1753: 489).

Although these were certainly not the first live green turtles seen in England, they were of sufficient rarity to gain newspaper comment. "Of all the improvements in the modern kitchen," said the *World* in an account of a London banquet at about this time, "there are none that can bear a comparison with the introduction of the turtle" (quoted in *Notes and Queries*, 1884 [6th ser., 9: 114-15]). But Dr. Samuel Johnson, in his *Dictionary* (1775), tersely defined "turtle" as a term "used among sailors and gluttons for a tortoise."

As English demand increased, vessels in the West Indian trade were provided with flat wooden tanks in which live turtles could be deck-loaded. Although they were fed grass and banana leaves on the journey, after arrival at the Leadenhall Street turtle tanks they still had to be fattened before reaching the tables of the well-to-do. The largest, which were not necessarily the best, were often destined for the royal palace. For the less affluent there already were substitutes. As early as 1808, Mrs. Elizabeth Raffald's *The Experienced English Housekeeper* was offering a recipe for "artificial" or mock turtle soup made from a calf's head.

Steamships greatly facilitated the movement of live turtles across the Atlantic. Imports of "preserved turtle" from Jamaica were initiated in 1841. The turtles had been taken by Cayman Islanders and, thus, by British subjects. By 1880 imports of "prepared turtle" were listed as 10,800 pounds. This was apparently the designation applied to the sun-dried meat and calipee that in late years had begun to place turtle soup, by one account, "within the reach of the general consumer." But it was to remain preeminently a prestige food. Mrs. Isabella Beeton called turtle soup "the most expensive soup brought to the table," with 1 guinea the standard price for a quart of it. Her widely read *Mrs. Beeton's Book of Household Management* states:

The price of live turtle ranges from 8d. to 2s. per pound, according to supply and demand. When live turtle is dear, many cooks use the tinned turtle, which is killed when caught, preserved by being put into hermetically sealed canisters, and so sent over to England. (Beeton 1861: 178-80)

The Queen of Soups

"Turtle soup from Painter's Ship & Turtle on Leadenhall Street," wrote one observer at the end of the nineteenth century, "is decidedly the best thing in the shape of soup that can be had in this, or perhaps any other country." Located there, he asserted, was the only "turtle artist" in Europe (Hayward 1899: 24-5). A French visitor described the establishment in 1904. A large pool of water contained upwards of 50 turtles awaiting "sacrifice." Alongside was the slaughter room and next to it the kitchen, where 10 to 12 men were occupied in making this national soup, which was sent out each day to the city, to the provinces, and even to foreign markets. It brought the exorbitant price of 1 guinea per liter for the regular soup and 25 shillings for the clear soup. A bowl of turtle soup served in the restaurant cost 3 shillings, the price including the glass of punch that followed it (Suzanne 1904: 15).

Blending and seasoning were of the greatest importance in soup making and called for much experience and "know-how." Some cooks insisted that the fins and steak or inside red muscle meat of the turtle were chiefly responsible for the flavor, and that genuine "real turtle soup" (the sort that jelled of its

own accord on cooling) was properly made only from a broth of turtle meat. To this, diced calipee was added as a relish. Others, especially in London, regarded beef stock as an essential basis of the soup, holding that without it turtle soup tended to lack character. In this they claimed the support of the famous *Maitre Chef* Auguste Escoffier.

Tinned turtle products entered midlatitude markets about the middle of the nineteenth century. Some of the first canneries were located within the tropics, close to sources of supply such as one at Pearl Lagoon, Nicaragua. Another in Key West, Florida, employing 10 vessels and 60 men for turtle gathering, was reported in 1880 to be turning out 200,000 cases a year. The "green fat," or calipee, was often tinned separately from the meat and soup. It was once customary to serve it as a side dish, a spoonful being added to the soup if desired. The largest shipments of tinned turtle products were to London, but New York was a substantial secondary market. Although turtle canneries operated from time to time in Jamaica, Nicaragua, Grand Cayman, Mexico, Australia, North Borneo, and Kenya, they were short-lived ventures. In later years, until forced cessation of the trade, the larger share of the green-turtle soup and red meat that went into cans was processed either in the New York area or in London.

The leading London soup maker, John Lusty, Ltd., "By Appointment Purveyors of Real Turtle Soup to the Royal Household since the Reign of Edward VII," had been in business near the London docks and Greenwich Naval Base since at least 1851 (long before the reign of Edward VII). Captains of Royal Navy ships returning from the West Indies or Ascension Island often brought back live green turtles and desired that they be made into soup for presentation to "My Lords of the Admiralty," who prized it as a great delicacy. The West Indies was the principal source of Lusty's supply.

After World War II, however, few live turtles were imported, the animals being slaughtered and refrigerated at the port of shipment. In its last years in business Lusty's supplies (mostly frozen) came from Kenya and the Seychelles. For Bender & Cassel, Ltd., the other major London producer, the Cayman Islands provided the supplies. It was at this time that a substantial demand for green turtle soup developed on the Continent. It became a standard feature on the menus of luxury restaurants, particularly in the larger cities and tourist centers of Germany, the Low Countries, and, to a lesser extent, France. There were canning operations in West Germany, France, Denmark, and Switzerland. For most of these concerns the Indian Ocean seems to have been the principal source of supply.

In the United States, the dominant company in the green-turtle and turtle-soup business was Moore & Company Soups, Inc., of Newark, New Jersey ("Ancora" brand), formerly located in Manhattan. It had begun making turtle soup in 1883. Initially, the turtles arrived at the port of New York deck-loaded on banana boats; later they were trucked, turned on their backs, from

Tampa or Key West. In the early 1960s, an average of two truckloads a week arrived at the Newark plant, where they were held in a pond until slaughtered. If frozen, the carcasses were placed in refrigerators.

The End of an Era

In 1962 I hazarded the guess that between 15,000 and 20,000 green turtles a year were finding their way, in one form or another, to the commercial markets of North America and Europe. The guess included those animals slaughtered on the Indian Ocean islands and elsewhere exclusively for their calipee. Although the aristocracy was consuming less turtle soup than in times past, the market had been immensely broadened. What was once reserved for the epicures of London and New York had become available on the shelves of quality grocery stores throughout Europe and America. The implications of such an expanding demand in the midlatitudes, coupled with the growing population of the tropical world, seemed ominous.

But today, with most nations of the world signatory parties to CITES, international commerce in turtles and turtle products has all but ceased, although the animals continue to be exploited on a limited scale by traditional coastal populations, and turtle meat or eggs may be marketed clandestinely. Egg collecting may be permitted under restraining community rules, as at the important olive ridley nesting beach at Ostional on the Pacific coast of Costa Rica or at the Sarawak Turtle Islands. But in Mexico, perhaps most conspicuously in sparsely settled Baja California, illegal poaching for commercial purposes remains widespread. In that country turtle eggs, as well as the penis of the reptile, are prized for their presumed aphrodisiac qualities.

At Mexico's Rancho Nuevo beach (on the Tamaulipas coast), famed for the massive synchronous nesting emergences (*arribadas*) of the rare Kemp's ridleys, eggs for a time were removed to a protected hatchery within hours of being laid to avoid depredation by coyotes, shore birds, and humans. Under a U.S.-Mexico cooperative "head start" program, hatchlings were transferred in large numbers to Texas beaches in the hope of imprinting them and establishing new rookeries (Rancho Nuevo is the lone known Kemp's ridley beach in the world). But the project did not produce the hoped-for results and was finally abandoned as futile (*Marine Turtle Newsletter* 1993: October, 63).

A film made in 1950 of the Tamaulipas *arribada* showed some 40,000 females storming the beach in daylight hours, but the same beach today supports fewer than 1,000 nesters in a season. Apparently the entire world stock of Kemp's ridleys has been severely depleted in the interim, many trapped and damaged in the trawls of commercial shrimpers (Cornelius 1990). Presently large *arribadas* of the vastly more abundant olive ridley turtle occur at one remaining Mexican West Coast beach and at others in Costa Rica and India. The phenomenon, one of the

most spectacular examples of mass activity in the animal kingdom, is especially characteristic of the two ridleys. At least three Mexican *arribada* sites had been eliminated by commercial fishermen before government restrictions began to be enforced. An estimated 2 million turtles were slaughtered on the West Coast of Mexico, more to meet demands for leather than for meat, in the five years leading up to 1969, before efforts were made to rationalize the harvest with catch quotas (Cornelius 1990: 54; *Marine Turtle Newsletter* 1991: April, 53). The India site, in Orissa State on the Bay of Bengal, of which the scientific world has only recently become aware, is reported to support *arribadas* of up to 200,000 ridleys, presumably in a single season. Despite the fact that much of the site is within a wildlife reserve, thousands of illegal takings occur offshore annually. The extent to which eggs are collected is unclear. The excessive harvest of eggs has been identified as one of the most important factors causing the decline of sea turtle populations.

In the Caribbean area, turtle deaths have increased with the intensified activities of shrimp trawlers in recent years (National Research Council 1990). Strandings of dead loggerheads on Florida beaches from this cause (13,000 in a recent year, with a small number of greens) have been sharply reduced by the Turtle Exclusionary Devices (TEDs) now required on the U.S. trawler fleet. Extension of their use to other nations is being urged. The ingestion of plastic bags, debris, and toxic substances such as petroleum further contributes to turtle mortality. Tourism and turtles, too, are on a collision course in several places. Coastal resort development and other forms of habitat encroachment deny nesting turtles their habitual nesting localities, as does the increasing artificial lighting along coastlines, to which the greens seem especially sensitive.

Gastronomy or Ecotourism?

Captive breeding programs for the more valued greens and hawksbills have been generally unsuccessful. The Cayman Islands Turtle Farm, established in 1968 as Mariculture, Ltd., for a time raised greens in tanks for export to commercial processors. In the face of CITES restrictions, it has been converted to a successful educational and tourist attraction. Production of meat (and perhaps eggs) continues on a reduced scale for local island consumption. The U.S. market has been closed since 1978, as, increasingly, are those of other countries (Wood 1991; Fosdick and Fosdick 1994). Provision for trade in farmed turtles continues to be sought, but the pressure against any trading in species whose wild populations are endangered or threatened is substantial. Sea turtle research, including the tagging and release of hatchlings and yearlings, has been an additional feature of the farm's activities. In recent years, farms have sought to be self-sustaining, independent of wild stocks of eggs or wild breed-

ing turtles. The animals are fed twice daily with a high-protein, pelletized Purina Turtle Chow. They are slaughtered at 4 years of age, at 20 to 30 kilograms.

If sea turtles have a future it seems likely to be in ecotourism rather than gastronomy. They are featured on the flag, seal, currency, and postage stamps of the Cayman Islands, reflecting their close association with the islands and their presumed emotional appeal to potential visitors. Elsewhere, too, as in Florida, in Costa Rica, on the Great Barrier Reef, and on some islands of the Aegean Sea, "turtle watching" is becoming a featured tourist attraction.

The groundswell of concern for the future of sea turtles that has put all but the Australian flatback on the endangered species lists has led to a surge in scientific research on the animals and on the causes and consequences of their decline (Bjorndal 1981; National Research Council 1990). Representative of this effort are the activities of the Marine Turtle Specialty Group of the International Union for the Conservation of Nature (IUCN) and the *Marine Turtle Newsletter*, a comprehensive quarterly now published in both English and Spanish by the Hubbs-Sea World Research Institute of San Diego, California. So are the symposia on Sea Turtle Biology and Conservation, which are annual workshops devoted to these questions.

Realization of the seriousness of the plight of sea turtles, underscored by the CITES trade restrictions, has led to the effective elimination of turtle steaks, turtle soup, and turtle eggs from the tables of all but a handful of tropical developing countries. We are losing, as a consequence, a palatable and nutritious marine food of unique cultural and historical significance. Turtle meat is remarkably lean, with 5 percent of calories in fat compared to 40 for most meats (James Stewart, M.D., personal communication). The eggs are comparable to chicken or duck eggs in protein content and are rich in vitamin A (Simoons 1991: 366-7). Both the world of gastronomy and the lives of many coastal populations of the tropics are being significantly impoverished as turtle and turtle products, a primary source of red meat and protein, are being forced from menus by the excessive pressures of commercialization as well as by the relentless increase in human numbers.

James J. Parsons

Bibliography

- Anon. 1753. *Gentleman's Magazine* 23: 489.
- Beeton, Isabella. 1861. *Mrs. Beeton's book of household management: A complete cookery book*. London.
- Bjorndal, Karen, ed. 1981. *Biology and conservation of sea turtles*. Washington, D.C.
- Carr, Archie. 1952. *The handbook of turtles*. Ithaca, N.Y.
- 1954. The passing of the fleet. *Bulletin of the Institute of Biological Science* 4: 17-19.
- 1973. "So excellent a fishe." New York.

- Cornelius, Stephen. 1990. *The sea turtles of Santa Rosa National Park*. San José, Costa Rica.
- Dampier, William. 1906. *Dampier's voyages (1679-1701)*. 2 vols., ed. John Masefield. New York.
- Fosdick, Peggy, and Sam Fosdick. 1994. *Last chance lost? Can and should farming save the green sea turtle? The story of Mariculture Ltd. - Cayman Turtle Farm*. York, Pa.
- Fryer, John. 1909. *A new account of East India and Persia 1672-81*. London.
- Goldsmith, Oliver. 1825. *A history of the earth and animated nature*. London.
- Hayward, Abraham. 1899. *The art of dining*. New York.
- Hendrickson, John. 1958. The green turtle (*Chelonia mydas* Linn.) in Malaya and Sarawak. *Proceedings of the Zoological Society of London*. 130 (Series A): 445-535.
- Labat, J. B. 1724. *Nouveaux voyages du Père Labat aux îles de l'Amérique*. 2 vols. The Hague.
- Marine Turtle Newsletter*. 1991-4. San Diego, Calif.
- National Research Council. 1990. *Decline of the sea turtles: Causes and consequences*. Washington, D.C.
- Nietschmann, Bernard. 1973. *Between land and water: The subsistence ecology of the Miskito Indians, eastern Nicaragua*. New York.
1979. *Caribbean edge: The coming of modern times to isolated people and wildlife*. New York.
1991. Miskito Coast protected area. *National Geographic Research and Exploration* 7: 232-4.
1993. Nicaragua's new alliance for Indian-Latin America. *Natural Geographic Research and Exploration* 9: 270-1.
- Parsons, James. 1962. *The green turtle and man*. Gainesville, Fla.
- Raffald, Elizabeth. 1808. *The experienced English housekeeper*. London.
- Rochefort, César de. 1606. *The history of the Caribby Islands*. 2 vols. London.
- Sauer, Carl. 1962. Seashore - primitive home of man? *Proceedings of the American Philosophical Society* 106: 41-7.
- Schaw, Janet. 1939. *Journal of a lady of quality, being a narrative of a journey from Scotland to the West Indies, North Carolina, and Portugal in the years 1774 to 1776*. New Haven, Conn.
- Simmonds, Peter. 1883. *Commercial products of the sea*. London.
1885. *Animal food resources of different nations*. London.
- Simoons, Frederick. 1991. *Food in China: A cultural and historical inquiry*. Boca Raton, Fla.
1994. *Eat not this flesh: Food avoidances in the Old World*. Madison, Wis.
- Suzanne, Alfred. 1904. *La cuisine et pâtisserie Anglaise et Américaine*. Paris.
- Walter, Richard. [1748] 1928. *Anson's voyage round the world*. London.
- Wood, Fern. 1991. Turtle culture. In *Production of aquatic animals*, ed. C. E. Nash, 225-34. Amsterdam.

Origin and Domestication

Sheep were domesticated on the flanks of the Taurus-Zagros Mountains, which run from southern Turkey to southern Iran. Within that arc is found the urial (*Ovis orientalis*), a wild sheep now generally regarded as the ancestor of the domesticated sheep. Early archaeological evidence of sheep under human control comes from Shanidar Cave and nearby Zaqi Chemi in Kurdistan. Sheep bones recovered in abundance at these two sites have been dated to between 8,000 and 9,000 years ago and contrast with other Neolithic sites close to the Mediterranean, where similar evidence of domesticated sheep is rare. However, accurate species identification has posed problems, for the bones of goats and sheep are often difficult to distinguish from one another. Therefore, some archaeological reports have grouped them together as "sheep/goat" or "caprine."

The domestication process that transformed *O. orientalis* into *O. aries* involved several key changes. The body size of the sheep was reduced from that of the urial. Diminution could have been accomplished over many generations by culling out larger, aggressive males as sires. Selection also occurred for hornlessness, but this process is not complete. Although many breeds of domesticated female (and some male) sheep typically have no horns, in other males the horns have only been reduced in size. Domesticated sheep also have a long tail as compared with the wild ancestor. The most significant physical transformation of the animal was the replacement of the hairy outercoat with wool fibers, which turned the sheep into much more than a food source. As early as 6,000 years ago, woolly sheep had differentiated from hairy sheep, and in ancient Mesopotamia, the raising of wool-bearing animals was a major activity in lowland areas. Selection for white-wooled animals explains the gradual dominance of that color.



Sheep

II.G.21 🐑 Sheep

Probably the earliest domesticated herd animal in the Old World, the sheep (*Ovis aries*) makes an unparalleled contribution of food and fiber. The great advantage of these small ruminants is their ability to digest the cellulose of wild grasses and coarse woody shrubs in their complex stomachs and convert it into usable products.

In spite of the many human uses of sheep, domestication may have been motivated by religion rather than economics. Urinals were animals of ritual significance, and to ensure a ready supply for sacrifice, humans may well have sought to tame and then breed them in captivity. At Catal Hüyük in Anatolia, the remains of sanctuaries from between the seventh and sixth millenniums B.C. depict ram heads.

The early use of sheep as sacrificial offerings went hand in hand with a long selection for qualities of fecklessness, timidity, and total dependency. Long after successful domestication, sheep sacrifices continued in religious ritual. The cultic use of sheep was known in ancient Egypt, Greece, Rome, and China. The Hebrew Bible made many allusions to ovine sacrifice, and from this, Christianity developed the idea of the "Lamb of God," a metaphor for Jesus as a sacrificial vessel for the sins of mankind. Unlike Christianity, Islam incorporated sacrifice as a literal requirement of the faith. Each year, during the festival of *Id al-Adha*, every male Muslim is enjoined to slaughter a domesticated animal, in most cases a sheep, as a sign of his submission to the will of God. Sheep accompanied the spread of Islam from its Arabian core, not only because of religious associations but also because the Arabs were, above all, pastoralists.

Ovine Diversity

Through isolation and/or mutation, sheep differentiated into almost 1,000 breeds, most with regional distributions. Some of these have now disappeared as more productive breeds have taken their places. The selection of sheep breeds in Western Europe has been influenced largely by the strong Western bias toward high productivity. During the eighteenth and nineteenth centuries, the breeds of Cheviot, Cotswold, Dorset, Hampshire, Leicester, Oxford, Romney, Shropshire, Suffolk, and Southdown all emerged in the British Isles, bred in some cases for their wool and in others primarily for their meat. The Rambouillet, a smooth-bodied wool breed, originated in France. In the United States, the breeds of Columbia, Debouillet, Montalde, Panama, and Targhee were developed (mostly in the twentieth century) through crossing of different breeds to adapt them to North American environmental conditions.

About 10 percent of classified breeds produce fine wool. Merino sheep are the outstanding source of high-quality wool and, as such, are the most important of all sheep breeds. The long fibers of Merino wool are turned into yarn used for worsted apparel. The history of Merino sheep in Spain is well documented after about A.D. 1500, but their origin is not. It seems most plausible, however, that this breed was brought from North Africa during the Moorish occupation of the Iberian Peninsula.

Genotypic and phenotypic diversity of sheep is greatest in the Near East, where these animals have

been part of human livelihoods longer than anywhere else in the world. Breeds of indigenous origin, well adapted to the local environmental conditions, still dominate there. Some breeds, such as the Awassi and Karaman, have fat tails – fat tails being a much-appreciated delicacy in places such as the Middle East; others have semifat tails (e.g., Chios sheep); still others have thin tails, for example, the Karayaka of Turkey. Fat-rumped breeds form yet another category. Near Eastern breeds have several kinds of fleece, or none at all. The good-quality carpet wool yielded by some breeds makes possible the manufacture of Oriental rugs, one of the world's magnificent art forms. Fur-sheep are another category; most famous are lamb pelts from Karakul sheep produced in the Middle East and central Asia, especially in Bokhara, where this breed originated. Many sheep of western Asia also have their variants in Africa (Epstein 1971).

In places where sheep are raised more for their subsistence value, numerous breeds may be represented in a single flock. For example, in northeastern Brazil, an owner may keep hair sheep and those with wool; polled sheep and those carrying horns; sheep with colors ranging from red-brown to white to black-pied; sheep with horizontal ears, but also those with lop ears; sheep with thin tails, but also those with semifat tails. These heterogeneous mixtures are the result of several introductions and free crossings over time. The first sheep were exported to Brazil in the sixteenth century from Portugal and included both the coarse-wooled Churro and the fine-wooled Merino. The Crioulo breed, coarse-wooled and horned, emerged from the Churro. Hair sheep were imported from the Caribbean and elsewhere, and in the twentieth century, sheep were brought from Italy to Brazil.

Keeping of Sheep

Sheep have been kept either as part of pastoral livelihoods or as an element of mixed agropastoralism. In the former, they consume grass and woody shrubs that are part of the natural vegetation. Sheep grazing on the coastal meadows (*prés salés*) of western France has established a standard for high-quality lamb production known throughout the culinary world. When integrated with a farming economy, sheep consume stubble in the fields and, in turn, contribute manure to renew the soil. Sheep are almost never kept in stalls or feedlots as cattle are. Intensive sheep production has lagged far behind that of cattle and pigs.

Where aridity has made agriculture too uncertain or impossible, sheep can survive on sparse wild grasses and woody shrubs. Sheep raising has traditionally been most profitable when natural vegetation on land owned by the community or the state reduced the cost of production. Often, however, the

need of such vegetation has involved seasonal movement to find proper forage at all times of the year. In much of the subhumid world, transhumance is the solution to providing livestock with what they need to survive. In the Iberian Peninsula, sheep raising became the most valuable avenue of land use after the introduction of the Merino. The powerful sheep-owners' organization, the *Mesta*, obtained priority in deploying their sheep over the countryside. Always in search of more and better pastures, sheep were driven north and south on the Spanish plateau along designated pathways called *cañadas*. The migration of millions of sharp-hoofed animals etched miniature canyons into the land that can still be seen as relict features in the landscape.

Traditional patterns of seasonal sheep movement continue to prevail in the Middle East, North Africa, and central Asia. In Europe, transhumance between high and low pastures is still practiced, but now sheep may be transported by rail and truck rather than on foot. Transhumance is also found in western North America and southern South America, but more in the form of commercial livestock strategy than cultural adaptation. In all forms of sheepherding, shepherds and dogs remain indispensable, for they protect their defenseless charges not only from predators but also from the sheep's own mimetic behavior and innate stupidity.

Sheep Products

Wool

Wool is the main product of commercial sheep raising such as is practiced in Australia, New Zealand, South Africa, Argentina, Uruguay, and North America. Most wool involved in world trade comes from these areas. Wool is also a major product for small farmers throughout the world. Sheep's wool has many advantages: It is resilient, which imparts to fabrics the ability to retain shape and resist wrinkling. Wool traps and retains heat-insulating air, but at the same time its low density permits the manufacture of lightweight fabrics. In addition, wool fiber takes dyestuffs well.

Meat

If all the world's sheep are taken into consideration, meat has been the primary objective of raising them. Fat (tallow) from sheep was once tremendously important for making candles until paraffin replaced it; in the Middle East and Africa, tallow continues to have considerable value as a substitute for cooking oils. Selection for fat tails is ancient. Herodotus mentioned sheep in which the tail constituted one-sixth of the total weight of the butchered animal. In some sheep, the tail is so large and heavy that it is an obstruction to the animal's mobility. Cases are known in which owners have constructed a little wheeled cart to relieve the weight of the tail and keep it from dragging on the ground.

Sheep meat is divided into mutton and lamb, with the latter deriving from an animal of up to 1 year of age and without permanent teeth. Mutton – the meat of a sheep older than 1 year – is the favorite meat of the Middle East and North Africa. Indeed, more than 50 percent of those regions' total meat requirements are satisfied by mutton. The sheep is the prestige animal of Islam, which ecology has encouraged, because the quality and quantity of forage in this part of the world favor it over cattle, which have much larger appetites and more stringent feed requirements in order to thrive. Goats are an alternative to sheep. Pigs are taboo.

Mutton is also culturally important in central Asia as a food, more so than in China, even though that country is the world's leading mutton producer. Australia, with its population of less than 20 million people, nevertheless produces four times as much sheep meat as the United States with 265 million people. Australians eat most of it but also export considerable quantities. The island peoples of Mauritius and Papua New Guinea, for example, derive a major part of their commercial meat supply from imported Australian mutton.

In Europe, sheep meat consumption is only one-eighth that of beef and veal and one-sixteenth that of pork. Among Europeans, the British have been especially fond of mutton, but their consumption of it has declined in recent decades in response to the availability of cheaper sources of protein such as poultry. Nonetheless, the United Kingdom still produces three times more sheep meat than the United States and also relies on mutton imports from Argentina.

The preferred sheep meat in many Western countries is lamb, which is more tender and subtle in flavor than mutton. A century ago, however, lambs were rarely marketed as a source of meat. One specialized variant of lamb is a milk-fed baby, which yields a succulent white flesh. In Mediterranean countries, suckling lamb is a much appreciated Easter delicacy. It is an old specialty (*abbaccho*) of the Roman Campagna, a favorite residential area of Rome in ancient times, and in Greece, milk-fed lamb is considered the height of gourmandise.

As already hinted, Americans eat little sheep meat of any kind. In 1985, less than 1 percent of the red meat consumed in the United States was lamb or mutton (USDA 1987). One explanation for this small amount is that sheep flesh was historically of poor quality because the only sheep slaughtered were those whose wool-bearing days were over. The nineteenth-century rise of the beef-centered meatpacking industry may also have played a role in marginalizing sheep meat. But most persuasive in explaining the weak pattern of sheep meat consumption are protein alternatives. Since 1963, the use of lamb and mutton in the United States has decreased more than 60 percent; during this time, pork, beef, and especially poultry have become relatively cheaper to purchase.

Another contributing factor is meat cuts: Mutton and lamb have a higher ratio of fat and bone than do beef or pork.

Today, what little sheep meat is consumed in the United States is as lamb. Part of that market is composed of immigrant populations, which reduces even more its consumption among mainstream American meat-eaters. The latter group may eat it mainly in restaurants specializing in Greek, Middle Eastern, and French cuisines. However, in some regions of the country, particularly the South and the Midwest, lamb is not even readily available in many supermarkets. Where found, in the meat counters of larger cities, lamb is typically at least as expensive as beef or pork, and usually more so.

Consumers who reject lamb for reasons other than price often state that they do not like its strong taste. The main factors controlling flavor in lamb meat are breed (Rambouillet lambs have a more intense flavor than Columbia lambs); sex (rams have a more intense flavor than wethers or ewes); and age and weight (flavor intensity varies inversely with these) (Crouse 1983).

Milk and Cheese

Milk has been a subsidiary product from sheep for millennia. Today, it is most important in the Middle East, where it is occasionally drunk fresh but is more commonly turned into yoghurt and cheese. Turkey, Iran, and Syria are major producers of dairy products from ewes, but most of these products are not commercialized beyond the local area. In Europe, the human use of sheep's milk is said to have increased when the Hundred Years' War killed off many cattle. Today, however, fresh sheep's milk is no longer consumed by most Europeans. The relative inefficiency of milking sheep, and their low productivity (150 pounds per lactation compared to 20,000 pounds for a cow), has encouraged the conversion of sheep milk into higher-value dairy products. In the Balkans, sheep's milk is made into yoghurt, but elsewhere in Europe, it is almost entirely made into cheese.

Europe's most famous sheep's-milk cheese is Roquefort – made distinctive by its veins of blue-green mold which develops on curd placed to cure in cool limestone caves. "Roquefort" is a controlled appellation (*appellation contrôlée*), which means that cheeses with this name can come only from a designated territory in the Cévennes region of France with the town of Roquefort as its center. However, the sheep's milk from which the curd is made comes from a much wider area that extends from the Massif Central to the Pyrenees and the island of Corsica. Almost all French sheep's milk produced is now devoted to making Roquefort cheese.

The manufacture of Roquefort is a holdout of tradition; other French cheeses once made with sheep's milk are now made from cow's milk. In Italy, sheep's-

milk cheese is called pecorino (*pecora* is the Italian for ewe); it can be either soft and fresh (ricotta pecorino) or hard (pecorino romano, or simply Romano, used in grated form on pasta). *Queijo da serra* ("mountain cheese") is a notable cheese from the mountainous interior of Portugal, a country that well into the twentieth century made most of its cheese from sheep's milk. Greece produces large amounts of feta, a sheep's-milk cheese salted and then preserved in a brine of milk, water, and salt.

The lands of the New World produce very little sheep's-milk cheese. About 90 percent of Latin American production comes from the Bolivian Altiplano near Oruro. The native cheese makers of *queso de Paria* are indigenous people whose ancestors were taught the art of making a soft, unripened cheese by the Spanish conquerors of the Andes. In the United States, sheep's-milk cheese is manufactured on a small scale in the upper Midwest and in California and Vermont. It is considered a gourmet item because sheep's milk commands four to five times the price of cow's milk. One advantage of sheep's milk is that it can be successfully frozen and stored. Hinkley, Minnesota, has a cheese-processing plant that derives its sheep's milk from a wide area.

Environmental Effects of Sheep

Shepherding has an ecological downside when considered historically over long periods. These ungulates have frequently overgrazed the land and brought about serious erosion, especially in areas where vegetation regenerates slowly. Sheep graze plants much closer to the ground than do cattle, and if uncontrolled, they can even, by consuming tree seedlings, denude an area to the point of preventing regeneration. Sheep contributed to the early formation of eroded landscapes in the Middle East and the Mediterranean region. Even a recently settled land such as Australia shows strong evidence of the sort of damage that intensive sheep raising can do. By 1854, New South Wales had a sheep population of over 12 million. It was an "ungulate irruption" that changed species composition of the vegetation, encouraged the introduction of noxious weeds, and degraded the soil. Sheep introduced into the New World have caused deterioration in the ranges of the western United States, Argentine Patagonia, and the Andean highlands. In Mexico, a case study of the Valle del Mezquital north of Mexico City has documented environmental changes wrought largely by intensive sheep grazing during the colonial period (Melville 1994).

World Production and Trade

More than 1 billion sheep are found in the world today, and they occur on all inhabited continents. Ovines are found in the hottest countries, such as

Somalia and Sudan, but there are also large herds in cold, windswept lands near Antarctica, such as Patagonia in southern Argentina and the Falklands (Malvinas), and on the fringes of the Arctic, as in the Faroe Islands and Iceland. Depending on the area, plains, plateaus, and mountains can all support sheep populations. Tolerance for hot, cold, dry, and wet conditions, along with an absence of any cultural prejudice against the animal and its multifaceted uses, account for its wide distribution.

Asia has about a third of the world's sheep, with especially large populations in China (mainly west and north), Iran, India, Kazakhstan, and Pakistan. Africa's sheep are found in all countries of that continent and, in the northern, western, and eastern parts, are often associated with nomadic, seminomadic, or transhumant groups. As a country (though not as a continent), Australia has, without question, the largest sheep population. Sheep are also the kingpin of agropastoralism in New Zealand, where the sheep-human ratio is among the highest in the world.

Europe (west and east combined) has about as many sheep as Australia. The superhumid United Kingdom and subhumid Spain have the largest ovine populations in the European Union. South America has more than five times as many sheep as North and Middle America taken together. Sheep are the most important domesticated animal in the high Andes of Peru, Bolivia, and Ecuador, where the cold climate and a homespun tradition make wool an especially valuable product. But 5 million sheep are also found in hot northeastern Brazil, where wool has little value. Much farther south, commercial wool production is important in temperate southern Brazil, Uruguay, Argentina, and Chile. The pastoralists of tiny Uruguay, which has about the same number of sheep as the United States, have long grazed sheep and cattle together.

Multispecies grazing in Uruguay contrasts with the conflictive tradition in western North America, where sheepmen and cattlemen pitted themselves against one another in a struggle for domination of the range. Since 1942, the rise of synthetic fibers, a shortage of skilled labor, and the increased cost of land have together forced a decline in the sheep industry in the United States. Nevertheless, entrepreneurial sheep raising on federally owned expanses still constitutes an important use of land in parts of the arid West.

Live sheep also find their way into world trade. Nigeria, Senegal, Kuwait, and especially Saudi Arabia import large numbers of them to be slaughtered in prescribed Islamic fashion. Movements of live sheep from Australia to Saudi Arabia are particularly large, enabling the faithful to satisfy their obligation of dispatching a sacrificial animal during the pilgrimage to Mecca. Normally, more than 1 million sheep are sacrificed each year for this purpose. Much of this slaughter, which wastes the meat, occurs in five abattoirs in Mina, near Mecca (Brooke 1987).

The long-term future of sheep raising in the world appears bright. Ruminant animals possess a keen advantage in being able to make use of arid or steep lands that cannot be cultivated. Sheep are productive, adaptable, and largely noncompetitive with humans. More than any other domesticate, their dual contributions of food and fiber give sheep an economic edge that spreads the risk of keeping them. And, finally, there are no cultural barriers to constrain their use.

Daniel W. Gade

Bibliography

- Brooke, Clarke. 1987. Sacred slaughter: The sacrificing of animals at the *hajj* and *Id al-Adha*. *Journal of Cultural Geography* 7: 67-88.
- Crouse, John D. 1983. The effects of breed, sex, slaughter weight and age on lamb flavor. *Food Technology* 72: 264-8.
- Epstein, H. 1971. *The origin of the domestic animals of Africa*. New York.
- FAO (Food and Agriculture Organization of the United Nations). 1995. *FAO yearbook*. Rome.
- Mason, I. L. 1980. Sheep and goat production in the drought polygon of northeast Brazil. *World Animal Review* 34: 23-8.
- Melville, Elinor E. K. 1994. *A plague of sheep: Environmental consequences of the conquest of Mexico*. New York.
- USDA (United States Department of Agriculture). 1987. *Food consumption, prices and expenditures*. Statistical Bulletin No. 749, USDA Economic Research Service. Washington, D.C.

II.G.22 🦃 Turkeys

The process of capture, taming, and eventual domestication of most animals is a difficult and lengthy process, often consisting of a trial-and-error approach. One notable exception was the domestication of the North American wild turkey, *Meleagris gallopavo*. The U.S. National Park Service archaeologist Jean Pinkley, while stationed at Mesa Verde National Park, put forth a logical scenario outlining the unique process of taming and domesticating the prehistoric pueblo turkeys of that area. Pinkley (1965) has pointed out that some domesticated animals apparently first exploited humans before becoming another of their agricultural conquests. The turkey is such an example.

The pueblo turkeys had become extinct in the Mesa Verde area by historic times, and the Park Service reintroduced breeding stock in 1944 (Pinkley 1965). This permitted observation of the wild turkeys and their relationship with the employees of the park. The



Turkey

turkeys were timid at first, but as they learned where food could be found – in this case, feeding stations that the government set out for small birds – they took over these sources. They also moved into warm roosting places available in the park's residential areas, and despite efforts to chase them away by tossing Fourth of July cherry bombs and firing guns into the air, the birds continued to congregate in and around park dwellings.

There is little reason to believe that prehistoric humans in Mesa Verde were not tormented by turkeys in much the same manner, and sooner or later, when it dawned on them that the birds could not be driven off or frightened away, the Pueblo Indians would out of despair have begun to corral them to protect crops and foodstores. At this point, recognition of the turkey as a source of food and materials for bone tools would presumably have been a logical next step.

That the turkeys were utilized over a wide area is evidenced by their numerous bones recovered from many archaeological excavations throughout the southwestern United States. All growth and age stages of the turkey – from eggs and poults on through old adults – are represented in these findings. Also of considerable interest are the pens that were constructed long ago in the pueblo region to keep the birds from straying from areas of confinement.

A number of turkey pens have been found associated with early pueblos. I recall walking down the sandy bottom of White Canyon in Utah in 1950 during the uranium boom and spotting a small two-room ruin high on the canyon wall overhead. After a tedious but not particularly dangerous climb, I reached the ledge to find the ruin virtually undisturbed. There were black-on-white bowl fragments lying about, along with stone grinding tools. On one end of the overhanging rock, where it joined with the smooth main cliff face, was a wedge-shaped turkey pen. It was constructed of cottonwood twigs, about the size of a finger in diameter, held together

by plant fibers. The gate had hinges made of leather thongs and had been secured with a tied leather loop. On the floor of the small pen was a thick layer of turkey droppings. As I untied the gate and heard the squeak of the leather thongs as it opened, I thought about the last person who had secured the empty cage, perhaps intending to return with more wild birds.

The determination of whether a turkey was domestic or wild on the basis of skeletal evidence alone is – unlike with many other domesticates – nearly impossible, and when such determinations are advanced they are “shaky” at best. But poultry experts at the Cooperative Extension Service at Clemson University in York, South Carolina, have arrived at an interesting conclusion regarding the color of turkey feathers that may help separate domestic from wild birds.

Color change in feathers is brought about, in part, by the presence of lysine, a biologically important basic amino acid ($C_6H_{14}N_2O_2$) and one of the building blocks of protein. In the case of turkeys, if lysine is not present in the diet in adequate amounts, their feathers lose some of their pigment. Turkeys generally obtain the needed amino acid from natural foods of whole protein, including insects, worms, and grubs. But many vegetable foods are low in lysine, and one of the poorest is corn. Because evidence indicates that corn was one of the major foods given to the Pueblo turkeys, and because confinement would have prevented the birds from foraging for other supplements, this diet may have contributed to the increase in white-tipped feathers that are found with domestic birds.

Two turkeys were and are present in the Americas. These are the ocellated turkey (*Meleagris “Agriocharis” ocellata*) of Central America and southern Mexico, and the turkey that is regarded as the Thanksgiving bird in the United States (*M. gallopavo*), which is found throughout much of North America.

The genus *Meleagris* and, perhaps, the species *gallopavo* are reported from the Pleistocene of North America. The birds are known from a Basketmaker site (see Table II.G.22.1) at Tseahaliso Cave, Canyon Del Muerto, Arizona, having an associated date of 3700 B.C. Turkeys were definitely domesticated by Pueblo I times (A.D. 750–900).

Table II.G.22.1. Southwestern chronology

A.D. 1700 to the present	Pueblo V
A.D. 1300–1700	Pueblo IV
A.D. 1100–1300	Pueblo III
A.D. 800/850–1100	Pueblo II
A.D. 750–900	Pueblo I
A.D. 450–750	Basketmaker III
A.D. 1–500	Basketmaker II
Pre-A.D. 1	Basketmaker I

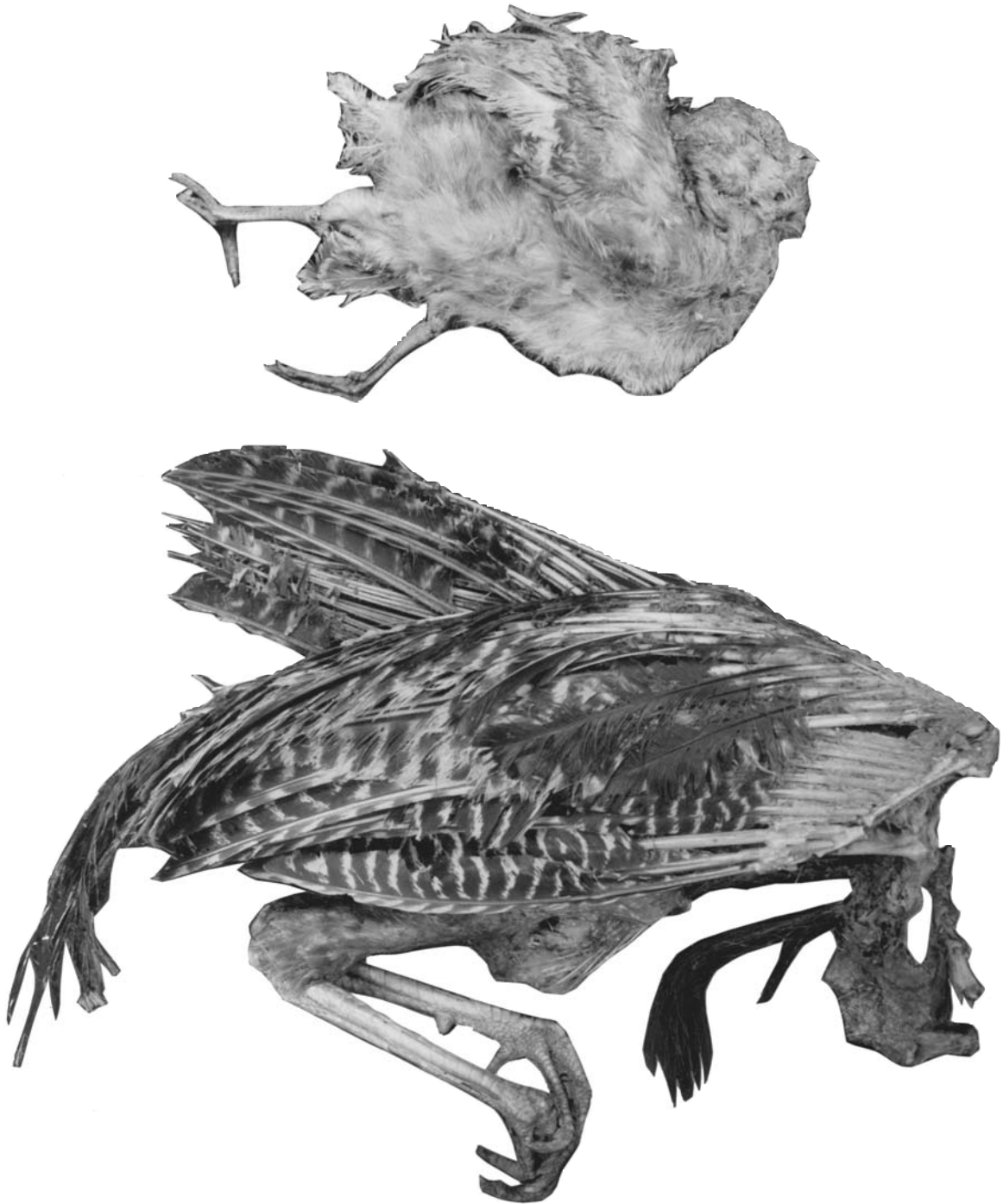


Figure II.G.22.1. Early pueblo domestic turkeys: top, a domestic turkey poult from a pueblo trash slope in Mesa Verde National Park, Colorado; bottom, natural mummy from an archaeological site in Canyon de Chelly National Monument, Arizona. (Photograph by John W. Olsen.)

Most historians agree that Christopher Columbus and his men were probably the first Europeans to see the turkey. During the fourth voyage of Columbus in 1502, his party landed at present-day Honduras, where friendly natives brought them food that included native birds the Spaniards called *gallinas de la tierra*, or “land chickens.” Subsequent Spanish visitors to the Mexican mainland also reported the turkey, such as Hernando Cortés, who saw them in the markets of the City of Mexico in 1519.

The ocellated turkey was long familiar to the Maya, and its bones are rather common discoveries in Maya sites in Yucatan and Guatemala. Indeed, the Maya referred to their part of the world as the land of the turkey and the deer. Yet there is little evidence indicating that the ocellated turkey was ever domesticated. Rather, it was probably a captive and not easily induced to breed in captivity.

The first European country to receive the turkey from the New World was Spain; Pedro Alonso Niño took some birds to that country in the early 1500s. The birds were established on Spanish poultry farms by 1530, were in Rome by 1525, were in France by 1538, and then spread rapidly to other parts of the Old World.

Ironically, the travels of the turkey came full circle when English settlers in Virginia brought the bird back to its home continent in 1584. The first turkeys to be brought from Europe to Massachusetts arrived in 1629. But contemporary accounts indicate that the turkey did not achieve its prominent place in the Thanksgiving festival in New England until the late eighteenth century.

This is not, however, to imply that turkeys were scarce in North America. Spanish explorers in the present-day southwestern United States, particularly those with Francisco Vásquez de Coronado from 1540 to 1542, supplemented their food supply with local animals encountered both in the wild and in Indian villages. Indeed, they frequently mentioned foods they encountered in the pueblos of Cibola and Cibuique, and Antonio Castañeda,¹ in his record of the Coronado expedition, noted that at the fortified village of Acoma (atop a large mesa in present-day New Mexico), the villagers gave the Spaniards presents of turkeys.

Coronado himself wrote of seeing domesticated turkeys in the southwestern pueblos. The Indians informed him, however, that the turkeys were not used as food but instead for their feathers that went to make robes (Winship 1896). Coronado, who did not take this statement seriously, claimed that the birds were an excellent source of food. Yet he also wrote that the Hopi pueblos kept both eagles and turkeys for their feathers. Feather-string robes similar to those reported in New Mexico have also been recovered from pre-Columbian Indian burials at several sites in Arizona.

Reports such as this one help explain why for many years archaeologists believed that turkeys were not eaten by the early Basketmakers and Pueblo Indians. But this theory also developed because of a number of complete natural turkey mummies – some buried with offerings of corn – that had been found (Figure II.G.22.1). This discovery was, however, before the thirteenth-century Mug House Ruin at Mesa Verde National Park was excavated (Rohn 1971) and over 1,074 turkey bones were recovered, representing from 183 to 815 individual birds (depending on the method of determining the minimum number of individuals).

Many bones at this site showed signs of butchering, and some were scorched or burned. In fact, turkeys appear to have been the most important source of meat for the Mug House occupants. The turkeys were housed mainly in the ruin’s “Room 46,” where numerous droppings found in most dry deposits, as well as remains of young birds or poults and some eggshell fragments, indicate that the birds were domesticated to a point at which they were reproducing. However, the limited number of eggshell fragments suggests that eggs were not an important food item (if, indeed, they were used for food at all). Grooved stones at the site may have been used as “anchors” to tether the birds and limit their wandering around the crowded pueblo (Rohn 1971).

Many of the turkey bones from Mug House, and other Pueblo sites, were made into awls, needles, and tubular beads. The bones selected to be “worked” were the humeri, femora, tarsometatarsi, and tibiotarsi. Even though turkey (and eagle) bones appear to be weaker in structure when compared with those of mammals, they are generally quite strong; thus, they could be worked and polished into implements.

A number of turkey humeri indicate that they were intentionally broken, or “captured,” and then allowed to grow back, leaving the birds unable to fly again. This method of “wing clipping” is also employed with macaws and is still used by bird collectors in Central America.

Depictions of turkeys are rather common in works of native art. The Hopi kivas at Awatovi Pueblo have identifiable turkeys on the walls. The Mimbres vessels of the Southwest have quite accurate portrayals of turkeys, as well as some other depictions with accurate turkey heads and beards but in combination with anatomical characteristics of reptiles and mammals, such as jackrabbit ears and deer limbs.

One of the more unusual artistic depictions of turkeys found in the southwestern United States is a rather large mural that was painted on the back wall of a cliff dwelling. The dwelling (Classic Pueblo, A.D. 1100–1300), in Arizona’s Navajo Tribal Park, is located some 60 to 70 feet above the floor of Little Canyon de Chelly. It is a site that is difficult to reach and, thus, was little known and virtually unvisited until the

1960s. The well-executed mural features three large turkeys roosting in a line.

Assemblages of turkey bones are preserved in many museum collections. Mug House Ruin, already mentioned, and Big Juniper House, also at Mesa Verde, have rather large collections of worked turkey bones. Seventeen bone awls representing the turkey, and one made from a golden eagle, were identified at Big Juniper House.

Worked turkey bones were also encountered in the following sites in prehistoric Arizona: Poncho House (4), Betatakin Pueblo (1), Kiet Siel Pueblo (77), Turkey Cave (3), Awatovi (1), and Wupaki Pueblo (1). Most of these turkey bones are from areas where, historically, the wild turkey is not known. Moreover, Lyndon Hargrave (1939), who did much of the identification of the bones, has pointed out that identifying single bones, particularly incomplete elements, can be less than accurate if turkeys, golden eagles, and sandhill cranes are all represented. Alden Miller (1932) has also reported on bird (including turkey) bones from archaeological sites in Arizona. These bones were from a dwelling site 35 miles north of Flagstaff. They were dated between A.D. 1000 and 1100, their age attested to by pottery types found with the turkey bones.

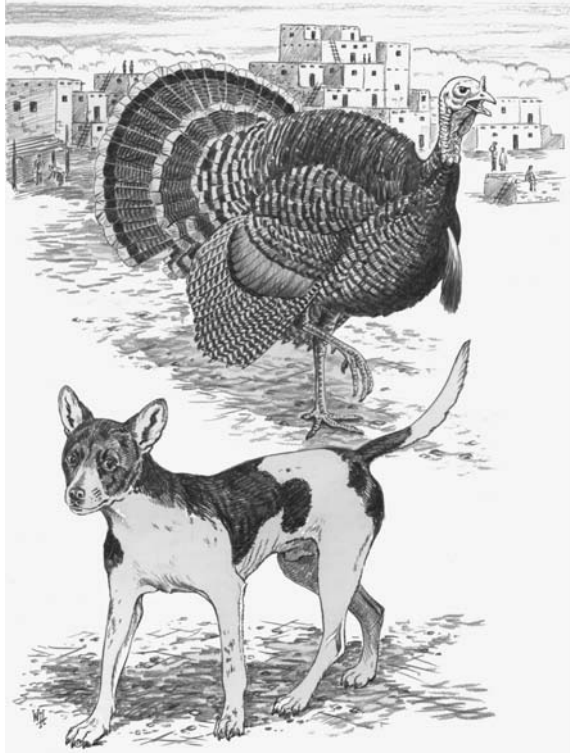


Figure II.G.22.2. Only two domestic animals greeted the first Europeans to visit the southwestern pueblos. They were the dog and the turkey. (Drawing by Wallace Hughes.)

Clearly, the turkey was prized for much more than just meat, and regardless of the degree to which *M. gallopavo* was considered as a food source in the Southwest, the evidence indicates that it was among the most sought-after animals over a wide area and a considerable span of time.

It is difficult to obtain reliable information on why turkeys were not fully accepted in many areas of the world where they were introduced. A. W. Schorger, in his book *The Wild Turkey; Its History and Domestication* (1966), devoted 36 pages to a discussion of the travels of the North American turkey to England, Norway, Germany, India, Portugal, and Africa, and throughout much of the Western Hemisphere, including Peru and Colombia. Nonetheless, North America seems to be the region where turkeys are most accepted, although no reason has ever been offered for such popularity.

The dog, *Canis familiaris*, and the turkey, *M. gallopavo*, were the only domestic animals present in North America during pre-Columbian times (Figure II.G.22.2). It is true that some workers credit the pre-Columbian Maya with domesticating the Muscovy duck, although my studies of the Maya fauna do not substantiate this hypothesis. In the Caribbean, Columbus observed the Carib Indians with penned (domesticated or merely captive?) muscovies. Obviously we still have much to learn about domesticated fowl in the early Americas.

Stanley J. Olsen

Notes

1. George Parker Winship (1896) presents the narrative of Castañeda in the original Spanish as well as an English translation. He also has compiled a most useful list of original Spanish works on the early Spanish explorers in the New World for the use of scholars who may wish to pursue this subject in greater detail.

Bibliography

- Hargrave, Lyndon L. 1939. Bird bones from abandoned Indian dwellings in Arizona and Utah. *The Condor* 41: 206-10.
- Jett, Stephen. 1977. *House of three turkeys*. Santa Barbara, Calif.
- Miller, Alden. 1932. Bird remains from Indian dwellings in Arizona. *The Condor* 34: 138-9.
- Pinkley, Jean M. 1965. The Pueblos and the turkey: Who domesticated whom? In *Memoirs of the Society for American Archaeology Number Nineteen*, ed. T. N. Campbell, *American Antiquity* 31: 70-2.
- Rohn, Arthur H. 1971. *Mug House*. U.S. National Park Service Archaeological Research Series No. 7-D. Washington, D.C.

- Schorger, A. W. 1966. *The wild turkey; its history and domestication*. Norman, Okla.
- Winship, George Parker. 1896. The Coronado expedition, 1540-1542. In *Fourteenth Annual Report of the Bureau of Ethnology, Smithsonian Institution, 1892-1893*, ed. J. W. Powell, 329-598. Washington, D.C.

II.G.23 Water Buffalo

When the first created man saw the animals that God had made, it is said that he presumptuously, over-rating his powers, asked that he too might be given the creative power to fashion others like them. God granted his request and man tried his prentice hand. But the result was the buffalo, and man seeing that it was not good, asked in disgust that the creative power might be taken back again from him for ever. The buffalo, however, remained as the only living handiwork of man. (Bradley-Birt 1910: 115)

Although of limited value as a clue to the origins of the domesticated water buffalo (*Bubalus bubalis*), this tale from India does reflect the rather low opinion of this bovine held by many, including, it seems, scientists who have shown relatively little interest in it. Considering the large size of its population, its widespread distribution, and its essential role in the economic lives of millions of people, especially in southern and eastern Asia, it is remarkable that so little is known about the water buffalo. Bovine admiration and scholarly attention have been reserved for those more highly regarded distant relatives of the buffalo, the taurine and zebu cattle.

Admittedly, some admirable efforts have been made within the last few decades to remedy this situation. Most of the work that has been done has focused on the present-day conditions and future potential of buffalo husbandry, with breeding, management, and productivity being of central concern.¹ Research on the cultural and historical aspects of the buffalo, however, has been very limited.² Certainly, in the discussion of animal domestication, the buffalo has been largely ignored.³

This chapter presents the results of a preliminary investigation of the original domestication and history of the water buffalo. A perusal of the available literature reveals a frustrating scarcity of information and very sketchy evidence. Consequently, interpretations must be highly speculative. At this stage in the research definitive answers are well beyond reach, and the best conclusions we can now offer are merely those that might guide future research.

Nomenclature

Although there is general agreement that the domesticated water buffalo represents a single species, its relationship to other bovines⁴ has been the subject of much disagreement. It received the name *Bos bubalis* from those naturalists who, beginning with Linnaeus in 1758, believed that bovines were all sufficiently similar to warrant classification under the same genus (*Bos*). Later taxonomists, using elaborate anatomical criteria, argued that bovine differences should be recognized by subgenera or genera status. In the many multiple genera and subgenera schemes that were proposed, a *Bubalus* genus or subgenus was always included for the classification of the water buffalo (and usually the other related "buffalo" - the *anoa*, *tamarau*, and African buffalo⁵).

The origin of the Latin appellation *bubalus* (or *bubalis*) is unknown. The learned ancients of Greece and Rome used the term, but seemingly in reference to an African antelope, most likely the bubal hartebeest (*Alcelaphus buselaphus*). From Roman through medieval times it was also applied to the European wild bison and the wild auroch (Buffon 1812: 304-8, 316-18; Pliny 1855: 262-3). When a Benedictine monk, Paul the Deacon, recorded the introduction of the domesticated Asian buffalo into Italy during the late sixth century A.D., he used the term *bubali* (Paul the Deacon 1907: 158-9; White 1974: 203-4). Thereafter, the accepted usage for *bubalus* was in reference to the Asian water buffalo.

The word "buffalo" was undoubtedly derived from the Latin *bubalus*. The Old English forms, *buffle* and *buff*, probably evolved through the French form *bufle*, but the present form seems to have entered English from the Portuguese *bufalo*. After some early European travelers in India erroneously used the word "buffalo" to refer to the zebu, the term "water buffalo" came into use to differentiate the true buffalo (Yule and Burnell 1903: 122; Murray et al. 1933: 1157). The confusion continued - as part of what Lynn White, Jr. (1974: 204), has called "the colonial transmission of antique and medieval perversity to North America" - when English speakers mistakenly used the word "buffalo" for the American bison. Use of the term "water buffalo" has been perpetuated as a means of avoiding this further confusion.

Biological and Ecological Characteristics

Appearance

The buffalo has a broad and massive body with short muscular legs ending in large splayed hoofs. Its shoulder height averages about 5 feet, and its length from nose to tail averages about 10 feet. The weight for adult buffalo varies from about 600 to 2,000 pounds. Although their massive bodies do not appear particularly graceful, they do reflect the animals' great strength (Figures II.G.23.1 and II.G.23.2).



Figure II.G.23.1 The domesticated water buffalo. (Photograph by the author.)



Figure II.G.23.2. Wild buffalo in Assam, with typical riparian and tall-grass habitat depicted. (Photo by Philippa Scott, from Pfeffer 1968: 156; courtesy of Chanticleer Press.)

Both sexes carry very distinctive transversely grooved and triangular cross-sectioned horns. These horns normally are crescent-shaped and incline upward, outward, and slightly backward following a single plane extending above and in the same plane as that of the forehead. Many variations in size and shape are found, from the short and tightly curled horns of the Murrah breed of northwestern India to the very long and nearly straight horns of the wild buffalo in Assam.

Shades of slate gray to black are the most common colors for the buffalo's skin and hair. Frequently the animal has light-colored chevrons under the jaw and on the chest and white or gray stockings. Although buffalo are born with a thick coat of long coarse hair, this hair generally becomes sparse as they age. In temperate or high-altitude locations, however, where cold weather is encountered, a heavy coat of hair is retained in the adult stage. Hair whorls are found in all breeds, and their individual distinctiveness has caused them to be used in some countries as a criterion for legal recognition and registration (Cockrill 1974a).

Adaptation and Habitat

Usually described as a denizen of the tropics, the buffalo is not capable of high heat tolerance. Its heat-regulating mechanisms are less efficient than those of European and Asian cattle, and under conditions of exposure to direct solar radiation, the buffalo exhibits a greater rise in body temperature, pulse, and respiration rates, and general discomfort. The buffalo is quicker, however, to show signs of comfort with a lowering of body temperature as a result of shade, artificial showers, or natural rain. As a means of cooling themselves and ridding themselves of insect pests, buffalo, unlike cattle, show a natural inclination to wallow.

The animal's preference for shade and semiaquatic conditions is at least partially explained by the color and structure of its skin. Like most tropical animals, the buffalo has black skin that absorbs ultraviolet light, thus preventing inflammation and damage to the deeper skin layers. But unlike most tropical mammals that have a coat of light-colored hair that reflects infrared waves, the buffalo has only a sparse coat of dark hair on its dark skin. Although this black surface facilitates the absorption of heat in direct sun, it is also advantageous in shade because of its high heat-radiating powers. The buffalo also suffers from the absence of an efficient perspiring mechanism. The number of sweat glands per unit area of buffalo skin is less than one-sixth that of Western cattle and an even lesser proportion of that of zebu.⁶

Thus, the buffalo is well adapted to a tropical environment provided that environment includes sufficient cooling systems in the form of shade and water. In the wild state these conditions are met with an

environment of high rainfall, heavy vegetation cover, and streams, lakes, swamps, or marshes for wallowing. Husbandry practices attempt to duplicate the conditions of the animal's natural habitat as much as possible by providing artificial shade, nearby water bodies for wallowing, and frequent splashing when needed.

The nature of the buffalo's feet, and its dietary habits, further indicates a semiaquatic, humid/tropical adaptation. The exceptional flexibility of its fetlock and pastern joints and its large splayed hoofs enable the animal to move easily in mud and water (Cockrill 1967: 124). Buffalo also reveal the remarkable ability to subsist entirely on those crude fibrous plants that form so much of tropical vegetation (especially around water courses), which other livestock will not touch. Experimentation has shown that compared to the zebu (its major economic rival in South Asia), the buffalo will consume more low-quality roughage, digest it more efficiently, and thereby maintain itself in better condition with greater strength and productivity (Whyte 1968: 218; Whyte and Mathur 1974: 555-6).

The "Swamp" and "River" Buffalo Types

Roderick MacGregor (1941) was probably the first to recognize a significant subdivision within the *B. bubalis* species. Based initially on his observations of differences in bathing habitat preferences, he distinguished the "Swamp buffalo," which was the native domesticated buffalo of East and Southeast Asia, from the "River buffalo," which was the type generally found in India and farther west.

A number of differences separate these two types. Swamp buffalo prefer mud-bottomed swamps and marshlands in contrast to the River buffalos' preference for rivers with firm bottoms and clear water. Swamp buffalo throughout their range seem to be morphologically very similar, with little specialization having occurred. They appear to be much closer to the wild buffalo than are the River buffalo. The latter, in contrast, show a great deal of regional variation, with many specialized breeds having been developed, especially for improved milk yield. The Swamp buffalo tends to have a shorter body, larger girth, shorter and thinner legs, and a shorter face than the River buffalo. Swamp buffalo horns are fairly uniform in size with the common widespread crescent shape; River buffalo manifest great variety in horn shapes, often with a spiraling curve close to the head.

Population, Distribution, and Uses

Before we venture into the distant past in search of clues to the buffalo's original domestication, it is valuable to view the present-day consequences of that domestication.⁷ As the result of many millennia of human-buffalo interaction, the buffalo currently

has an extensive geographical range and is involved in a great variety of husbandry contexts. Recent estimates place the total world population of domesticated buffalo at about 142 million (Table II.G.23.1). Although these animals are reported from 31 countries (Map II.G.23.1), a high degree of concentration characterizes this distribution: Ninety-six percent reside in the countries of South, Southeast, and East Asia, with 80 percent in just the three countries of India, Pakistan, and China. India leads all other countries with 54 percent of the world total (about 77 million buffalo). Relatively small numbers are found in Southwest Asia, southern and Eastern Europe, and South America. The only countries within this peripheral zone that possess fairly significant numbers (over 1 million) are Egypt and Brazil.

East and Southeast Asia

In the area east of India, the Swamp buffalo dominates and is essentially a single-purpose animal, contributing primarily to the traction needs of the agricultural system. It is in the wet rice fields of southern China and Southeast Asia that the buffalo distinguishes itself as the preferred draft animal. The humid climate, the mud, and the water suit the animal's constitution and habits. No other domesticated animal can equal the buffalo's ability to carry out the plowing, harrowing, and puddling tasks of wet rice cultivation. Its considerable weight causes it to sink deep in the mud; the flexibility of its fetlock and pastern joints, its large splayed hoofs, coarse limbs, and bulky body allow it to maintain a balanced traction in mud and water; its enormous strength enables it to pull the plow deep into the soil.

In China, the cattle to buffalo ratio is about 4:1, with the buffalo distribution limited almost entirely to the rice-growing area in the southeast. Cattle are preferred in the drier and more temperate northern and western regions (Phillips, Johnson, and Moyer 1945: 62-6; Epstein 1969: 26-32; Cockrill 1976). But in mainland and island Southeast Asia, where wet rice cultivation dominates the agricultural economies, the buffalo becomes more important. The average cattle to buffalo ratio throughout Southeast Asia is about 2:1. Three countries - Laos, Brunei Darus, and the Philippines - actually possess more buffalo than cattle.

Eastern and southeastern Asia is located beyond the traditional limits of milking and milk use in Asia (Simoons 1970). The peoples in this area generally do not include milk or milk products in their diet and exhibit high incidences of primary adult lactose intolerance.⁸ A limited market for buffalo milk, however, has been found in some cities and in the small Indian and European communities scattered throughout the region.

Buffalo are rarely raised specifically for meat pro-

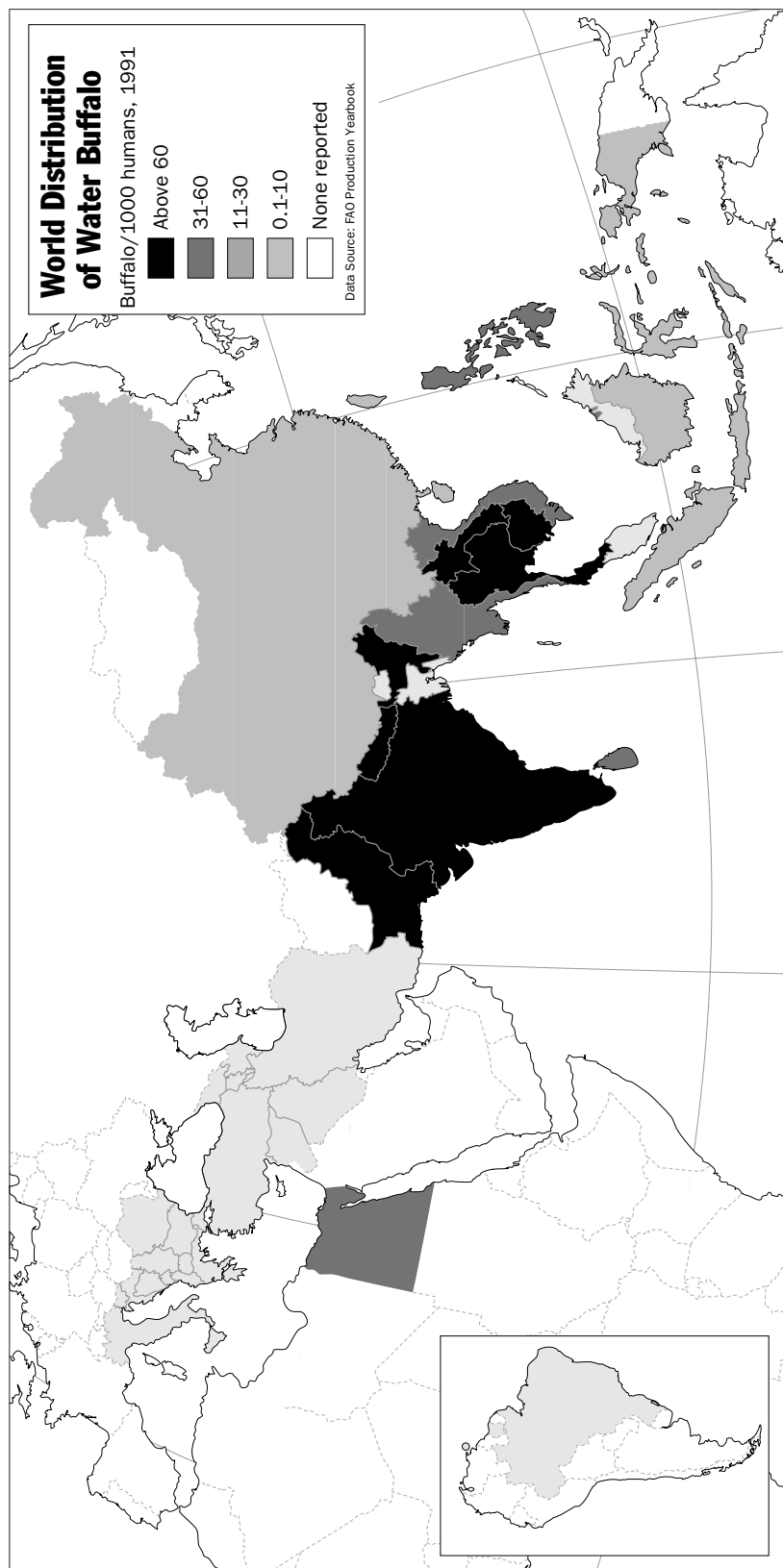
Table II.G.23.1. *Population of water buffalo (based on national livestock censuses and FAO estimates for 1991)*

Region/Country	Population	Buffalo per 1,000 people
South Asia		
Bangladesh	810,000 ^a	6.8
Bhutan	4,000 ^a	2.6
India	77,000,000	88.4
Nepal	3,101,000 ^a	158.2
Pakistan	15,031,000	118.7
Sri Lanka	981,000 ^a	56.3
Southeast Asia		
Brunei Darus	10,000 ^a	36.5
Indonesia	3,500,000 ^a	18.6
Kampuchia	760,000 ^a	18.7
Laos	1,100,000 ^a	257.9
Malaysia	190,000 ^a	10.4
Myanmar	2,080,000 ^a	48.9
Philippines	2,710,000	42.4
Thailand	4,743,000	84.0
Vietnam	371,000	43.0
East Asia		
China (incl. Taiwan)	21,635,000	18.7
Southwest Asia and Africa		
Egypt	2,550,000 ^a	47.6
Iran	300,000	5.4
Iraq	110,000 ^a	5.6
Syria	1,000 ^a	0.1
Turkey	371,000	6.5
Europe and the former Soviet Union		
Albania	2,000 ^a	0.6
Bulgaria	26,000	2.9
Greece	1,000 ^a	0.1
Italy	112,000	1.9
Romania	180,000 ^a	7.8
Yugoslavia (former)	20,000 ^a	0.8
Former Soviet Union	434,000 ^a	1.5
South America and the Caribbean		
Brazil	1,490,000	9.7
Trinidad and Tobago	9,000 ^a	6.9
Suriname	1,000 ^a	2.3
World Total	142,189,000	26.4

^aFAO estimate.

Source: FAO (1992), pp. 191-3.

duction or for any other purpose. Retired work animals, after having provided 10 to 20 years of service, are either slaughtered or allowed to die a natural death, at which time they contribute meat, horns, and hides to the local economy. While alive, the buffalo produces a very important by-product for the farmer - manure, a valuable alternative or addition to scarce and expensive chemical fertilizers.



Map II.G.23.1. World distribution of water buffalo.

South Asia

The buffalo is found in its greatest concentration in South Asia and there exhibits its greatest biological diversity and attains its greatest utility as a dairy animal. The distributions of the River and Swamp buffalo meet, and in some places overlap, in South Asia, and, although River buffalo predominate, Swamp buffalo are found in the Assam and Tamil Nadu states of India, and in Bangladesh, Sri Lanka, and Nepal (Porter 1991: 290-1). Selective breeding of River buffalo, primarily in India and Pakistan, has resulted in the formation of several distinct dairy breeds and even a few draft breeds. Probably between 15 and 20 percent of the buffalo in this region can be identified as representatives of defined breeds; the rest are nondescript local varieties (Hoffpauir 1974: 141-51).

As in eastern Asia, the buffalo in South Asia performs an important service as a work animal, especially in the wet rice fields, but this role is subordinate to that of milk production. In India, for example, only about 15 percent of all buffalo are classified as work animals, and they account for only about 10 percent of the total bovine workforce. The more versatile zebu bullock is the generally preferred work animal.

Even though the zebu outnumbers the buffalo in India by about 2.6 to 1, and buffalo account for only about one-third of the bovine population that is milked, over 60 percent of the milk produced comes from buffalo (Hoffpauir 1982: 223). In Pakistan and Nepal, fully two-thirds of the milk produced comes from buffalo (Cockrill 1974b: 605; Khan 1974: 612). Nature (presumably aided by selective breeding) has provided the buffalo with certain milk-producing abilities that are distinctly superior to those of the zebu - namely, the ability to yield more milk (usually between two and three times more), the ability to produce milk with higher butterfat content (between 6.5 and 8.5 percent, compared with 3.0 to 5.0 percent in zebu milk), and the ability to produce this milk while subsisting on poor-quality fodder (Hoffpauir 1977).

In Hindu India the slaughtering of the buffalo and the eating of its flesh is not generally practiced. This is probably explained by the general *ahimsa* (the Hindu/Jain/Buddhist principle of noninjury to living things) and vegetarian proclivities of the Indian culture, rather than by any special status similar to that of the sacred zebu. The Moslems in Pakistan and Bangladesh have no prohibition on the slaughtering of buffalo, but buffalo meat is unpopular because of the poor quality that is marketed. Most Pakistanis and Bangladeshis prefer beef (Cockrill 1974b: 535; Khan 1974: 611).

The population of buffalo in Bangladesh is far less than might be expected, given the predominance of wet rice cultivation throughout this Ganges-Brahmaputra delta area. In fact, it appears that the disasters this country has suffered in recent decades (typhoons, floods, earthquakes, and warfare) have

decimated the buffalo population. Most of the buffalo still there are of the Swamp type and are used mainly for work. Small numbers of dairy buffalo have been imported, but milk production remains low (Cockrill 1974b: 533-5).

The adaptability of the buffalo to high altitudes is clearly demonstrated in Nepal, where this tropical animal is maintained in village settings as high as 9,000 feet and is taken for summer pasturage as high as 15,000 feet. In Nepal, the buffalo is truly a multipurpose animal: It provides work power, milk, manure, and meat. In the tropical lowlands of the *terai* (the northern extension of the Indo-Gangetic plain in southern Nepal), it is the preferred draft animal for wet rice cultivation; it is seldom used for work elsewhere. In the lower and middle altitudes, the buffalo is the primary source of milk. Buffalo have been introduced into Nepal from India for the purpose of improving milk production. Whether the native Nepalese buffalo are of the Swamp or River type has yet to be determined. In many mountainous locations, the buffalo is used exclusively for the production of manure, an extremely valuable resource given the low fertility of the mountain soils. Nepalese law permits the killing of male buffalo, and considerable numbers are slaughtered for ritual sacrifice and for the urban meat market (Rouse 1970, 2: 905-6; Cockrill 1974b: 603-9; Epstein 1977: 38-46; Hoffpauir 1978: 238-41).

In many ways Sri Lanka follows the typical Southeast Asian pattern of buffalo use. Its indigenous buffalo is of the Swamp type and serves primarily as a work animal in the rice-growing areas. Some buffalo milk is produced and sold in the form of curd, but generally buffalo owners are not interested in milk production. Frederick Simoons (1970: 561-3) has commented on the survival of a nonmilking attitude among the Sinhalese, the dominant group in Sri Lanka. With a population that is predominantly Buddhist and Hindu, Sri Lanka has a limited demand for meat. Only old, sick, or infertile buffalo are allowed by law to be slaughtered; so the buffalo meat that reaches the markets tends to be tough and without taste (Cockrill 1974b: 629-35; Porter 1991: 291).

Southwest Asia and Egypt

There are few environments in the dry southwestern end of Asia and northern Africa that are suitable for the buffalo. Consequently, those in Iran, Iraq, Syria, Turkey, and Egypt account for only about 2.4 percent of the world's total. The only country in this region where buffalo make a significant economic contribution is Egypt. For the Southwest Asian countries, the cattle to buffalo ratio is 26:1; for Egypt it is 1.4:1.

All of the buffalo in this region appear to be of River buffalo ancestry, having been introduced from South Asia during the first millennium A.D. According to historical records, Sassanians brought domesticated buffalo and their herders from Sind in Pakistan to the marshes of the lower Tigris and Euphrates Delta dur-

ing the fifth century A.D. Buffalo were established in the Jordan Valley by the eighth century and in Anatolia by the ninth century (White 1974: 204-5). They were not known in ancient Egypt and do not seem to have appeared in the Nile Valley until about the ninth century A.D. (Epstein 1971: 567-8; Cockrill 1984: 57).

The 2.5 million buffalo in the Nile Valley of Egypt today are fully utilized for milk, work, and meat. Egyptians have thoroughly researched their productivity and have developed an efficient husbandry system. Buffalo are highly valued as milk animals and produce far more milk than cattle. Although cattle are more important than buffalo for draft purposes, the latter are used for plowing and harrowing in rice fields, for raising water from wells for irrigation, and for threshing. John E. Rouse (1970: 621-2) reported that only buffalo cows are used for work. The common practice is for male calves to be slaughtered at a very young age for veal, which brings a high market price. Female buffalo are also slaughtered for their meat, but only if they are old, infertile, or injured (Epstein I: 1971: 564; El-Itriby 1974; Porter 1991: 299-300).⁹

Not surprisingly, given the water needs of this bovine, the major concentrations of buffalo in Southwest Asia are in swampy areas and along coasts. The buffalo in Southwest Asia is esteemed only by small numbers of peasant farmers and herders who recognize the animal's unique adaptability to environments that are unable to support other livestock. Throughout this area the buffalo is valued primarily as a milk animal¹⁰ and used very little for work, except in Turkey, where in some areas buffalo are employed in plowing and road haulage. Buffalo meat is merely a by-product from old retired animals (Rouse 1970, 2: 794, 804, 854; Cockrill 1974b: 510-32; Porter 1991: 297, 298).

Europe and Transcaucasia

The existence of the buffalo in the temperate latitudes of Italy, Greece, the Balkan peninsula, and Transcaucasia in the former Soviet Union attests to the animal's ability to tolerate cold winter temperatures. But their numbers are small - the total population being only about 775,000 - and care must be taken to keep these animals warm during the winter; cold winds and sudden drops in temperature can cause fatal illnesses, including pneumonia (National Research Council 1981: 45-6).

All European and Transcaucasian buffalo are of the River type. Historical accounts are sketchy but indicate that domesticated buffalo may have entered Europe by a route north of the Black Sea, from Persia to southern Russia to the Danube Valley and, finally, to Italy by the end of the sixth century A.D. Another likely route could have been from Turkey to Greece, the Balkans, and Italy. They were established in Transylvania by the eleventh century and were numerous in the area around Rome by the twelfth century (Bökönyi 1974: 151; White 1974: 203-6).

The discontinuous and spotty distribution pattern that characterizes the buffalo's existence in Southwest Asia is also found in Europe. In Greece, for example, 90 percent of the buffalo are found in Macedonia and Thrace. In Romania there are two concentrations, one in the central regions near Cluj and Fagaras, and another in the Danube Valley of the south. The Italian buffalo are found mainly in the southern part of the country. The only buffalo in the former Soviet Union are in the Transcaucasus region, where they are most numerous in the Kura Valley and along the Caspian coast in Azerbaijan.

Throughout most of its European and Transcaucasian distribution, the buffalo has served as a triple-purpose animal. Its use as a work animal, originally of primary importance, has declined in recent times with the mechanization of agriculture, especially in Italy, Greece, and Bulgaria. Meat production is growing in importance, although buffalo meat still comes mostly from old animals whose usefulness for work and milk production has been exhausted. The buffalo is highly valued for its rich milk, which is often made into yoghurt and cheese. In Italy, where probably the best buffalo in Europe are found, the emphasis has long been on milk production, with the highly popular mozzarella cheese the major product (Cockrill 1974b: 708, 731-5, 748-54; Polikhronov 1974; Salerno 1974; Porter 1991: 297-9).

Tropical America

Introductions of buffalo into various Caribbean and South American countries during the past 100 years have met with mixed results. Those into Bolivia, Colombia, French Guiana, Guyana, Peru, and Venezuela were apparently unsuccessful, and either the animals have entirely disappeared or their numbers have become too insignificant to be reported in recent censuses. Possibly a thousand buffalo can still be found in Surinam, derived from an 1895 introduction of Swamp buffalo from Southeast Asia, via French Guiana, to work on the sugar plantations. They are used today in logging operations and for meat production. Indian River buffalo were brought into Trinidad and Tobago during the early 1900s, also for the sugar plantations. A sizable population of about 9,000 animals has developed, and they are highly valued for their work power and meat (Cockrill 1974b: 676, 692-7, 705-7; Mahadevan 1974; Porter 1991: 300-2).

The most successful establishment of buffalo husbandry in the Western Hemisphere has been in Brazil. Numerous introductions of both Swamp buffalo (from eastern and southern Asia) and River buffalo (from India and Italy) have occurred over the last century. The nearly 2.5 million Brazilian buffalo today are concentrated in two areas, with about 75 percent in the Amazon Basin and the rest in and around the states of São Paulo and Minas Gerais. The buffalo are used primarily for milk in the latter areas, with meat production secondary although growing in importance. In the Amazon area the buffalo has found a paradise, with

ideal environmental conditions. On Marajó Island, large herds of free-ranging buffalo thrive on the tropical grasses and aquatic plants. Buffalo meat and milk command a ready market in the towns of Amazonia, and the animal is also used for road haulage and riding (Gade 1970; Cockrill 1974b: 677–91; Porter 1991: 301).

The Wild Ancestor

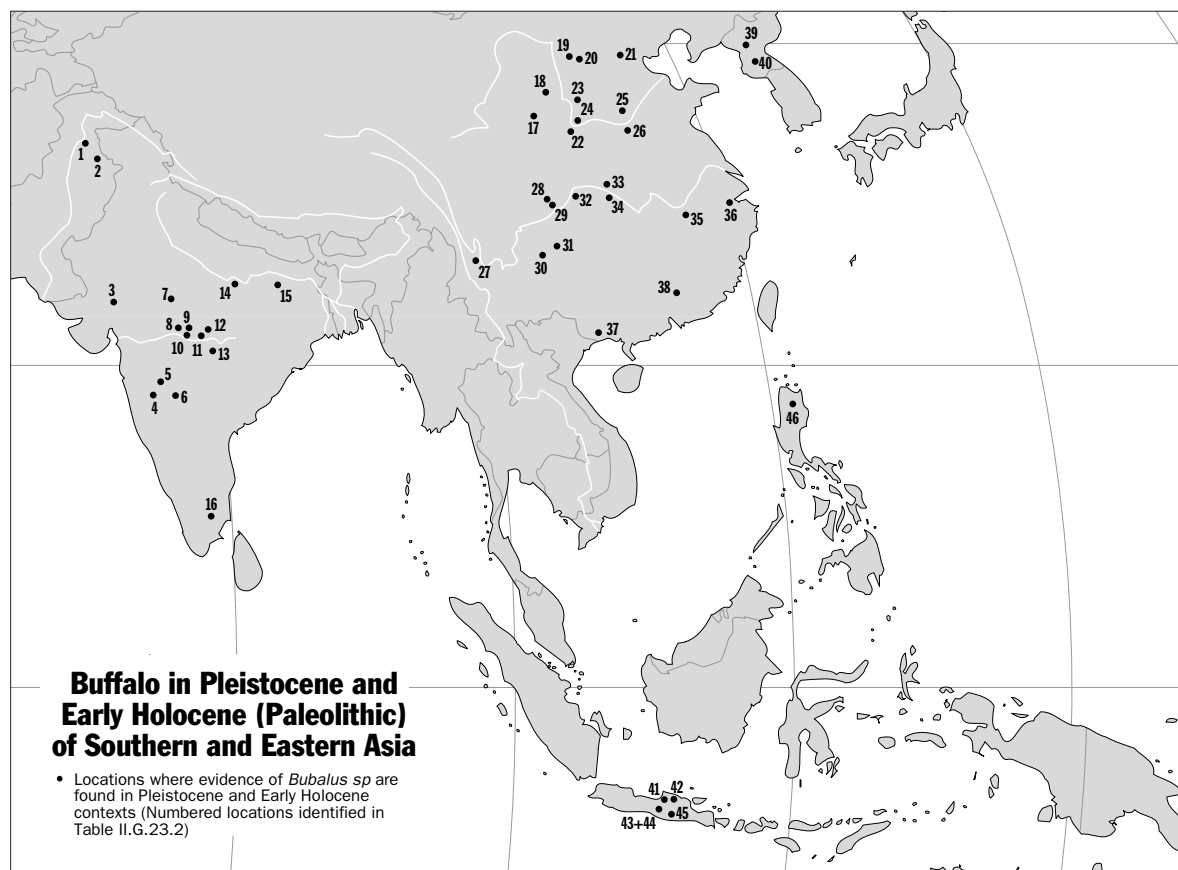
Unlike the progenitors of many other domesticated animals, the wild form from which the domesticated buffalo originated is not only known but still survives in a few locations. Wild buffalo differ from the domesticated form chiefly in their greater size, larger horns, and generally more robust and sleek appearance.

The Plio-Pleistocene Buffalo

The evolutionary history of the wild buffalo appears to have begun about 3.5 million years ago in the northwestern corner of the Indian subcontinent. The earliest fossil to show uniquely *Bubalus* characteristics is *Proamphibos* from the upper Dhokpathan Formation (Middle Pliocene) of the Siwalik Hills (Groves 1981: 270, 276).

Within South Asia, when the conditions in the northwest became unfavorable on account of Pleistocene glaciation, the Siwalik fauna, including the buffalo, apparently migrated southward and eastward. The Middle Pleistocene form *Bubalus palaeindicus* has been discovered in the Potwar Plain of northern Pakistan as well as in the Narbada Valley of peninsular India. By the Upper Pleistocene the buffalo was found as far south as Tamil Nadu and as far east as Bihar. As revealed by bone finds and cave paintings¹¹ over a wide range of the Indian subcontinent, the buffalo was undoubtedly hunted by Paleolithic and Mesolithic peoples throughout the Pleistocene and into the Holocene (Map II.G.23.2 and Table II.G.23.2).

The buffalo appeared in China as early as the Middle Pleistocene (possibly between 1 million and 500,000 years ago), and its remains are often found associated with early hominids/humans in Paleolithic sites. Although most *Bubalus* finds from Chinese sites are reported as being of indeterminate species, some distinct species have been identified. A dozen Pleistocene sites in South China (south of the Qin Ling Mountains) have yielded buffalo remains.



Map II.G.23.2. Buffalo in Pleistocene and Early Holocene (Paleolithic) of southern and eastern Asia.

Table II.G.23.2. *Evidence of the existence of Bubalus sp. in the Pleistocene and Early Holocene (Paleolithic) of southern and eastern Asia (identification of numbers on Map II.G.23.2)*

Location	Time	Source
India/Pakistan		
1. Campbellpore	Lower/Middle Pleistocene	Terra and Teilhard de Chardin 1936: 796
2. Potwar Plain	Middle Pleistocene	Movius 1944:21
3. Langhnaj	Mesolithic	Sankalia 1946: 148, 314; Fairservis 1971: 95, 100
4. Ghod Valley	Upper Pleistocene	Badam 1985: 413
5. Upper Godavari Valley	Upper Pleistocene	Badam 1984: 756
6. Upper Manjar Valley	Upper Pleistocene	Badam 1984: 757
7. Chibbarnala	(unknown)	Mathpal 1984: 19
8. Itar Pahar	(unknown)	Mathpal 1984: 18
9. Bhopal area	Mesolithic	Mathpal 1984: 16
10. Bhimbetka	Mesolithic	Mathpal 1984: 105-6
11. Adamgarh	Mesolithic	Allchin and Allchin 1968: 83, 256; Sankalia 1978: 87
12. Putli Karar	(unknown)	Mathpal 1984: 17
13. Narbada Valley (near Narsinghpur)	Middle Pleistocene	Terra and Teilhard de Chardin 1936: 820; Wheeler 1959: 43-44
14. Belan and Seoti Valleys	Middle/Upper Pleistocene	Sankalia 1977: 66; Badam 1984: 755
15. Paimar Valley	Upper Pleistocene	Badam 1984: 755
16. Ariyalur region	Upper Pleistocene	Badam 1984: 765
China		
17. Qing-yang	Upper Pleistocene	Aigner 1981: 258
18. Sjara-osso-gol	Upper Pleistocene	Aigner 1981: 83, 84, 248, 303; Defen and Chunhua 1985: 275, 285; Jia and Huang 1985: 212
19. Shuo-hsien	Middle/Upper Pleistocene	Aigner 1981: 225
20. Shiyu	Upper Pleistocene	Pearson 1983: 120; Wu and Wang 1985: 44
21. Zhoukoudian (Localities 1, 7, 9, 13)	Middle Pleistocene	Aigner 1981: 81, 108, 116, 148, 302, 311-14; Defen and Chunhua 1985: 272, 282
22. Dali	Early Upper Pleistocene	Defen and Chunhua 1985: 284; Qiu 1985: 207
23. Dingcun	Early Upper Pleistocene	Aigner 1981: 84, 303; Defen and Chunhua 1985: 284
24. K'o-ho	Middle Pleistocene	Aigner 1981: 82, 178, 302; Chang 1986: 48
25. Hsiao-nan-hai	Upper Pleistocene	Aigner 1981: 89, 303
26. Lingching	Upper Pleistocene	Aigner 1981: 96
27. Lichiang district (Locality 6003)	Upper Pleistocene	Aigner 1981: 59, 295
28. Tungnan	Upper Pleistocene	Aigner 1981: 295
29. Koloshan	Middle Pleistocene	Aigner 1981: 295
30. Chihchin	Middle Pleistocene	Aigner 1981: 295
31. Guanyindong	Middle Pleistocene	Defen and Chunhua 1985: 283
32. Yenchingkou	Middle Pleistocene	Aigner 1981: 295
33. Chihchiang cave	Middle Pleistocene	Aigner 1981: 295
34. Changyang	Middle/Upper Pleistocene	Aigner 1974: 31; Defen and Chunhua 1985: 286
35. Yungshan cave	Middle/Upper Pleistocene	Aigner 1981: 71
36. Jiande	Upper Pleistocene	Defen and Chunhua 1985: 286
37. Ya-p'u-shan and Ma-lang-ch'uan	Early Holocene	Aigner 1981: 64; Chang 1986: 86
38. Northern Kwangtung	Middle Pleistocene	Aigner 1981: 295
Korea		
39. Tokchon cave	Upper Pleistocene	Pokee 1984: 880
40. Sangwon Komunmoru cave	Middle Pleistocene	Pokee 1984: 880, 882
Java		
41. Sampung cave	Early Holocene	Medway 1972: 81; Bellwood 1985: 37, 220
42. Ngandong	Upper Pleistocene	Movius 1944: 86; Koenigswald 1951: 219; Medway 1972: 80; Jacob 1978: 17
43. Djetis	Lower Middle Pleistocene	Medway 1972: 79
44. Trinil (same site as #43, but at higher level)	Middle Pleistocene	Medway 1972: 79; Jacob 1978: 16
45. Patjitan	Late Middle Pleistocene or Early Upper Pleistocene	Movius 1944: 90
Philippines		
46. Cagayan Valley	Middle Pleistocene	Fox 1978: 79; Fox 1979: 230

The Pleistocene buffalo that appear in North China are interpreted as immigrants from the south, and their existence as far north as 40° north latitude attests to their adaptability to cooler and drier environments. The buffalo apparently extended its range eastward from North China into Korea during the Middle Pleistocene. *Bubalus* has been uncovered at two cave sites, Sangwon Komunmoru and Tokchon. As in North China, this presence appears to represent a remarkable adaptation to what must have been a rather harsh Pleistocene environment.

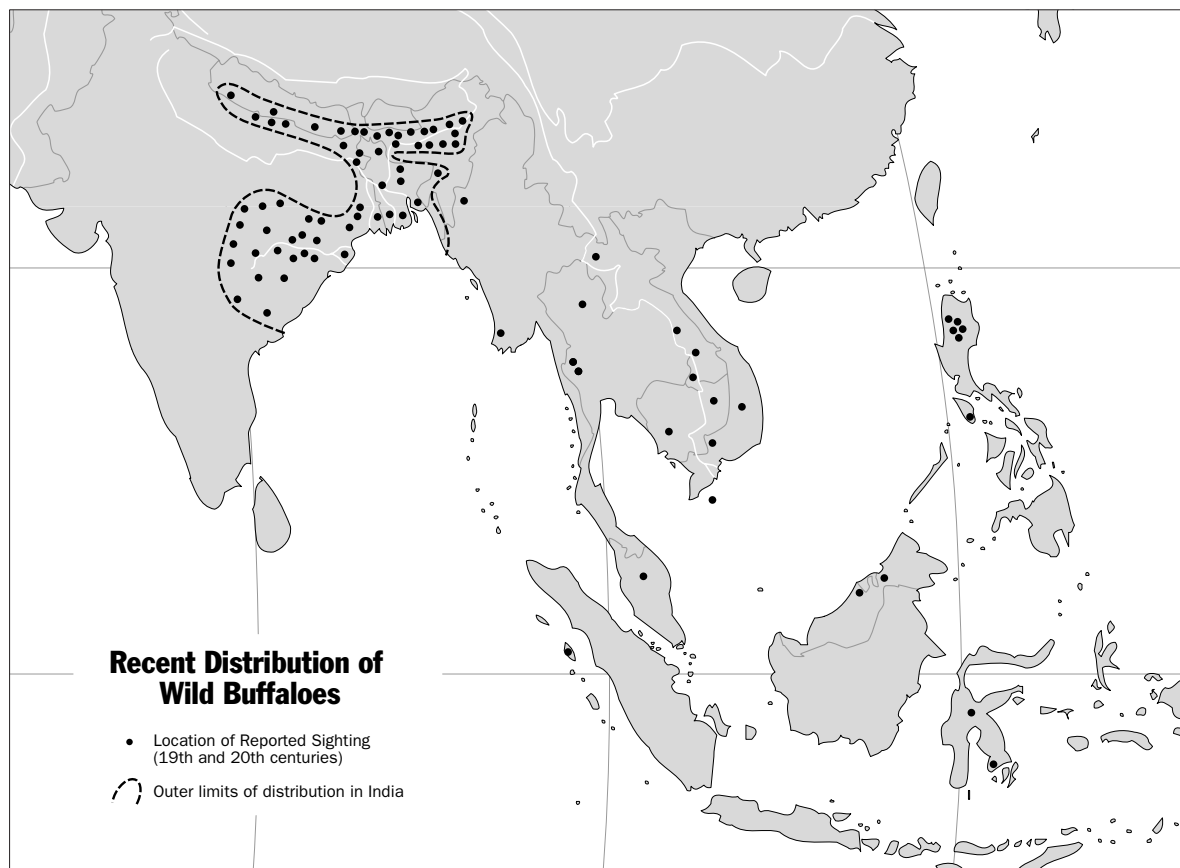
Buffalo ventured out onto the land masses of the Sunda shelf probably as early as they reached China. Although no Pleistocene buffalo have been firmly identified as yet from mainland Southeast Asia, they must have existed there, or at least passed through on their way to Sundaland. A very large form of a now extinct buffalo *Bubalus paleo kerabau* inhabited Java from the early Middle Pleistocene to the early Holocene. The only other indication of the buffalo in island Southeast Asia¹² comes from the discovery of *Bubalus* in association with the Liwanian Flake Tool Tradition in the Cagayan Valley on the Philippine island of Luzon. Robert Fox (1978: 82, 1979: 229–30)

has suggested that Paleolithic hunters, along with the buffalo and other typical Middle Pleistocene mammals, migrated from South China into the Philippines across the then-existing land connections via Taiwan.

A long-distance westward migration of Pleistocene buffalo is evidenced by the finding of *Bubalus* at three sites in Europe (not shown on Map II.G.23.2).¹³ A form called *Bubalus murrensis* is known from the two sites at Steinheim and Schöneck in Germany, both dated from the Mindel-Riss interglacial, about 250,000 years ago, and Riss-Würm interglacial site of possibly 100,000 years ago in the Peneios Valley of Greece. The buffalo seem to have been able to exist in Europe only during the interglacial periods and had disappeared by the beginning of the last (Würm) glaciation (Zeuner 1963: 246; Kurtén 1968: 187; Bökönyi 1974: 149–50).

Holocene Survivals

The wild buffalo probably maintained a broad range throughout southern, eastern, and southeastern Asia well into the Holocene. Evidence of *B. bubalis* has been reported from many archaeological sites in this area, dating from between 10,000 and 2,000 years ago. But whether these findings reflect locations before or



Map II.G.23.3. Recent distribution of wild buffaloes.

after domestication is, in many cases, difficult to say because of the morphological similarity between wild and domesticated buffalo. (The significance of these Neolithic and early Metal Age findings to the question of earliest domestication is discussed later in this chapter.) Consequently, we do not have a clear picture of the distribution of the wild buffalo several thousand years ago when it probably was first domesticated.

Reported sightings during the last two centuries indicate that the wild buffalo's range has generally retreated to east of the 80° meridian in southern and southeastern Asia (Map II.G.23.3). In South Asia, the wild buffalo are known in the *terai* of Nepal, the plains of Bengal and Assam, and lowland areas in Orissa, Madhya Pradesh, and Andhra Pradesh. They are described as generally inhabiting moist lowland riverine tracts where dense and high vegetation cover, usually tall grass jungles, and nearby swamps are available.

The distribution of the wild form of a domesticated species has often proved a valuable indicator of that animal's original place of domestication. This, unfortunately, is not the case with the buffalo. All that can be concluded from the available evidence is that the wild buffalo's distribution has responded to changing environmental circumstances that have occurred during and since the Pleistocene. As conditions under which the buffalo can exist have appeared and disappeared, its range has changed accordingly. Generally, the animal's extremely broad range during the Pleistocene, especially during the interglacial stages, has been shrinking throughout the Holocene. Its recent limited distribution is best seen as merely its last remaining refuge area and not necessarily as indicative of an ancestral hearth of domestication.

Domestication

Archaeological Evidence

The most convincing and satisfying evidence for the place and time of domestication of an animal species would be well-preserved fossilized bones that can be clearly identified as to species and domesticated status. Unfortunately, such evidence is relatively rare for the buffalo because the warm and moist environments in which buffalo usually exist make the chances of good skeletal preservation poor. Even when *B. bubalis* bones can be identified, as they have in many Neolithic and Metal Age sites (Map II.G.23.4 and Table II.G.23.3), the ability to distinguish between wild and domesticated forms is hindered by the absence of clear morphological differences. Consequently, fossil evidence must be very cautiously interpreted.¹⁴

South Asia. The earliest indication of a possibly domesticated buffalo in South Asia comes from the Indus Valley civilization (about 2300–1750 B.C.). The evidence, collected from the two sites of Mohenjo-daro and Harappa, consists of only a few fossils (teeth, horn cores, and bones¹⁵), depictions on a small number of seal-amulets

(Figure II.G.23.3), and two small figurines.¹⁶ Many authorities have accepted the notion that the Indus Valley buffalo was domesticated, and some have even suggested that these finds represent the original domestication.¹⁷ Such a conclusion, however, seems to be based solely on the fossils' structural resemblance to the domesticated form and on the fact that a few seals show the buffalo standing in front of an object that has been interpreted as a feeding trough (Figure II.G.23.3).

Evidence that the buffalo existed in the Indus Valley only in the wild state is far more convincing. There are four scenes from seals and prism-amulets that seem to depict the buffalo being hunted. In three scenes (Figure II.G.23.3) a man is shown with a foot on the animal's head, one hand grasping a horn and the other hand holding a spear about to be thrust into the beast. In another scene the buffalo appears to have attacked a number of people who are sprawled around the ground.

When the sex of the animal can be detected, it is always male. As with the depictions of all other bovines, the Indus Valley seals seem to emphasize the power and strength of the bulls rather than the services derived from them or the cows. There are no scenes of the buffalo being milked or pulling a plow or cart, and the so-called food trough also appears on other seals in front of tigers and rhinoceroses. In fact, the troughs are not shown with those animals that were most likely domesticated, such as sheep, goats, elephants, and zebu.¹⁸ Sir John Marshall (1931, 1: 70), the original excavator of Mohenjo-daro, suggested that the troughs represent offerings of food to worshiped animals, both in the wild and in captivity. Others have interpreted the troughs as cult objects or symbols connected to sacrificial ritual (Hiltebeitel 1978: 779).

Interpreting religious content from artifactual remains is, at best, a dangerous endeavor. With the Indus Valley culture in mind, Sir Mortimer Wheeler (1968: 108) has reminded us of "the notorious incapacity of material symbols to represent the true content and affinity of a religion or belief" and of "the indivisibility of religious and secular concepts in ancient times." Much of the attempt to decipher the religion of the Indus Valley people has focused on the interpretation of one particular seal scene, called the "Proto-Siva" scene, in which the buffalo plays a prominent role (Figure II.G.23.3). The significance of the buffalo and other animals shown around the seated godlike figure and the meaning of the buffalo-horn headdress worn by the seated figure are but two of the many concerns scholars have had about this picture. Some writers have assigned an important role to the buffalo in the religious symbolism and ritual life of these people (Sastri 1957: 6–13; Sullivan 1964; Srinivasan 1975–6; Hiltebeitel 1978). If true, the implications could be profound for the question of domestication, suggesting, for example, a ceremonial motivation.

Within the riparian environment of the Indus River and its tributaries, the wild buffalo could have found



Figure II.G.23.3. Depictions of water buffalo on seal-amulets from Mohenjo-daro. (From Mackay 1937–38.)

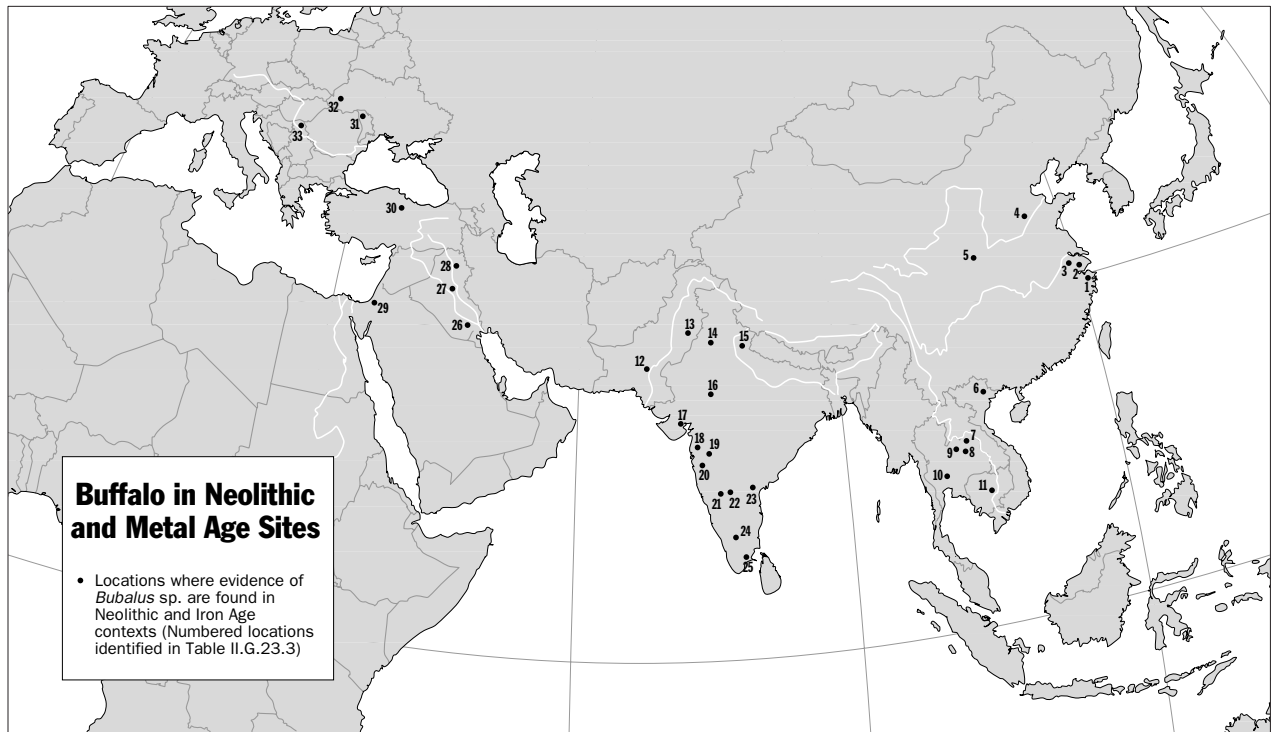
the marshy and grassy habitats that it prefers. The safest conclusion that can be reached at this point is that the wild buffalo existed in the Indus Valley during the third millennium B.C.; it was possibly hunted for its meat and may even have played some role in the religious life of the people. But the case for the existence of the domesticated buffalo is weak and unconvincing.¹⁹

The buffalo appears in later archaeological sites in India (see Map II.G.23.4 and Table II.G.23.3), but the domesticated status continues to be indeterminable. Small numbers of buffalo bones have been uncovered in Neolithic and Copper Age sites, dated from the second and first millennia B.C., south and east of the Indus Valley. But the majority of the bones at these sites belong to the domesticated zebu (*Bos indicus*), and the number of buffalo bones is very small, sometimes no more than a single bone at a site. Some of the bones are charred and have cut marks on them, a clear indication that the animals were cooked and

eaten. An occasional successful hunt of the wild buffalo could easily explain these remains.

The north Indian site of Hastinapura (1100 B.C. to A.D. 300) is said to reflect the expansion of the Indo-Aryan peoples from their original homeland in the Punjab region into the Ganges Valley of Uttar Pradesh. Although the fossil evidence (12 buffalo bones, some charred and with cut marks) is no more convincing than at other sites, it is probably safe to say that these were domesticated buffalo.

We can draw this conclusion because, although archaeological evidence fails us, literary evidence, at least for early historic northern India, throws some light on the question. The oldest literary work, the *Rig Veda*, which is believed to reflect Indo-Aryan culture as early as 1500 B.C., makes reference to “buffaloes yoked in fours,” the eating of buffalo flesh by the gods, and the slaughtering of from 100 to 300 buffalo at a time as sacrificial offerings to the god Indra (Griffith 1963, 1: 489, 575, 2: 123, 133, 226).



Map II.G.23.4. Buffalo in Neolithic and Metal Age sites.

Table II.G.23.3. Evidence of the existence of buffalo at Neolithic and Metal Age sites (identification of numbers on Map II.G.23.4)

Location	Time	Source
China		
1. Ho-mu-tu culture	6500 B.P.	Meacham 1983: 159; Pearson 1983: 124; Bellwood 1985: 219–20; Chang 1986: 211; Zhao and Wu 1986–7: 30–1; Smith 1995: 126, 140
2. Ma-chia-pang culture (aka Lake Ta'i-ha culture and Ch'in-lien-kang culture)	6500 B.P.	Pearson 1983: 131; Chang 1986: 201; Smith 1995: 125
3. Liang-chu culture	6000–4000 B.P.	Chang 1986: 254
4. Anyang	3800–3200 B.P.	Sowerby 1935: 234–7; Creel 1937: 76, 80; Coon 1955: 331; Chang 1980: 31–3, 138–43
5. Shaanxi phase of Lung-shan culture (aka K'o-hsing-chung)	5000–4000 B.P.	Ho 1975: 96; Chang 1986: 279
Southeast Asia		
6. Bo-lum (Bacson culture)	10,000 B.P.	Gorman 1971: 308; Meacham 1977: 423, 437
7. Ban Chiang	3600 B.P.	Higham and Kijngam 1979: 213–16; Higham and Kijngam 1985: 420–2
8. Ban Tong	probably after 4000 B.P.	Higham and Kijngam 1979: 218–19
9. Don Khlang	probably after 4000 B.P.	Higham and Kijngam 1979: 220
10. Chansen	2000 B.P. (Metal Age)	Wetherill 1972: 44, 45; Higham 1977: 404
11. Somron-Seng	possibly 2500 B.P. (Late Neolithic)	Boriskovskii 1970–1: 243
India/Pakistan		
12. Mohenjo-daro	4300–3750 B.P.	Marshall 1931, 1: v, 70–2, 386, 659, 666–8; Prashad 1936: 8–9, 43–6; Mackay 1937–8, 1: 45, 284, 292–3, 298, 311, 330, 336–9, 358–62, 670, 671; Vats 1940: 323; Fairservis 1971: 274, 275, 276, 277
13. Harappa	4300–3750 B.P.	(same as no. 12)
14. Kalibangan	Pre-Harappan/Harappan	Agrawal 1971: 29
15. Hastinapura	3100–1700 B.P.	Lal 1954–5: 107–8, 115–16; Allchin and Allchin 1968: 265
16. Ahir	4000–2500 B.P.	Sankalia 1977: 104
17. Rangpur	4000–2800 B.P.	Rao 1962–3: 155, 157; Allchin and Allchin 1968: 264; Fairservis 1971: 307

(continued)

Table II.G.23.3. (Continued)

Location	Time	Source
India/Pakistan		
18. Nasik	3000–2500 B.P.	George 1955: 142–3; Krishnaswami 1960: 39
19. Nevasa	3250 B.P.	Agrawal 1971: 42
20. Inamgaon	3600–2700 B.P.	Clason 1975: 6; Badam 1985: 414
21. Piklihal	Deccan Neolithic (about 4000–2400 B.P.)	Fairservis 1971: 325
22. Maski	Chalcolithic (about 2600 B.P.)	Thapar 1957: 13, 121–9; Krishnaswami 1960: 41; Agrawal 1971: 43; Fairservis 1971: 328
23. Nagarjunakonda	3000 B.P.	Soundara Rajan 1958: 109–10; Krishnaswami 1960: 42
24. Nilgiri Hills	2000 B.P.	Foote 1901: 22–47; Rea 1915: v, vi; W. A. Noble, personal communication
25. Adichanallur	2000 B.P.	Rea 1915: v, 4
Southwest Asia		
26. Ur	Pre-Akkadian (prior to 4500 B.P.)	Zeuner 1963: 249–50
27. Accad	4500–4100 B.P.	Duerst 1908: 361; Zeuner 1963: 249
28. Grai Resh	Uruk culture period (5500–4900 B.P.)	Lloyd 1947: 6–7
29. Tell el Hesi	3400–3300 B.P.	Bliss 1894: 191
30. Boghazköy	3200–2700 B.P. (Post-Hittite layers)	Bökönyi 1974: 151
Europe		
31. Frumusica	Neolithic	Bökönyi 1974: 150
32. Bukovina	Neolithic	Bökönyi 1974: 150
33. Csoka	Neolithic	Bökönyi 1974: 150

The *Yajur Veda*, dating from the early part of the first millennium B.C., mentions buffalo “reared for purposes of cultivation” (Chand 1959: 234). Although no earlier references exist, it appears that at least by the time of the writing of the *Sutras* (about 800–300 B.C.), the use of the buffalo for milk was well established (Bühler 1882: 73, 1896: 63; Jolly 1900: 166, 167).

This record probably represents the earliest use of the buffalo as a dairy animal anywhere. Although literal interpretations of this ancient literature are not always possible, it would seem safe at this point to conclude that the buffalo was a domesticated animal in northern India at least 3,000 years ago. For the wetter eastern parts of India, where environmental conditions suit the buffalo particularly well, archaeological and literary evidence is notably scarce. Consequently, the antiquity of the domesticated buffalo in eastern India cannot, as yet, be clearly shown.

Southwest Asia. Mesopotamia presents us with a situation very similar to that of the Indus Valley. The buffalo was undoubtedly well known in ancient Mesopotamia, but its domesticated status is disputed. One site in northern Iraq (Grai Resh), dating from the Uruk culture period (about 3500 B.C.), has yielded a horn core of a buffalo. Most of the remaining evidence is representational, coming from depictions on cylinder seals from the third millennium B.C. Arguments in favor of the buffalo’s domesticated status rely mainly on scenes showing the animal in the company of humans (or possibly gods or mythical heroes), in some cases being watered and fed by the human figures (Figure II.G.23.4). Some supporters of this argument propose

that the buffalo was actually domesticated by the Sumerians or Akkadians, and others suggest that the animal was imported from the Indus Valley already in a domesticated state.²⁰ These seal scenes might well reflect, as suggested by Henri Frankfort (1939: 85–94), religious and mythical symbolism, or maybe just artistic imagination or descriptive design. Wild buffalo probably existed at that time in the swamps of the Tigris and Euphrates Delta and could easily have inspired the seal depictions (Duerst 1908: 360–2; Ward 1910: 414–5; Bodenheimer 1960: 50, 102; White 1974: 202).

Whether wild or domesticated, buffalo seem to have disappeared from Mesopotamia possibly by the end of the first millennium B.C., although they may have survived longer in other parts of Southwest Asia. Frederick Bliss (1894) reported that fossil water buffalo teeth, dated between 1700 and 500 B.C., were found at Tell el Hesi, near Gaza in ancient Palestine. Buffalo bones were also found in post-Hittite layers of Boghazköy in north central Turkey, dated between the twelfth and seventh centuries B.C. (Bökönyi 1974: 151). A Sassanian seal depiction of a buffalo (illustrated in Bivar 1969: 83, plate 15 EM1) indicates that the buffalo still existed (or at least was still remembered by artists) in northern Iraq or Iran during the first half of the first millennium A.D.

China. Because of the close association of the domesticated buffalo with wet rice cultivation throughout southern and eastern Asia, some authors have suggested the possibility that the buffalo was originally domesticated as a work animal in the rice-growing area. They usually point to southern China or the

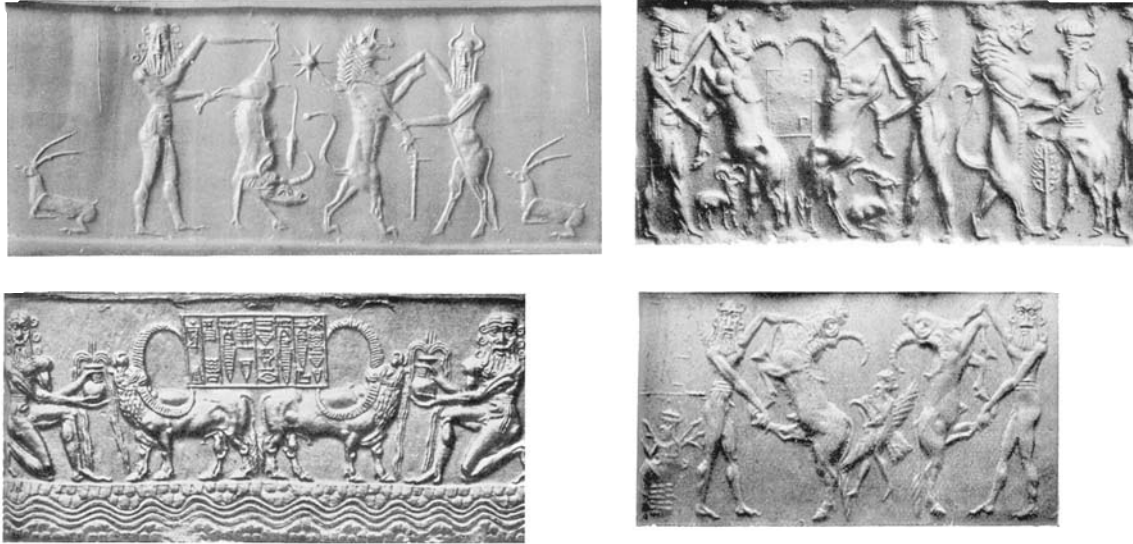


Figure II.G.23.4. Depictions of water buffalo on cylinder seals from Mesopotamia (Akkadian dynasty, about 2300–2100 B.C.). (From Frankfort 1939.)

mainland of southeastern Asia (Zeuner 1963: 251; Bökönyi 1974: 150–1; Clutton-Brock 1981: 140). Until recently there was no archaeological evidence to support this contention. But that has all changed with the astonishing archaeological finds in China during and since the 1970s.

Between about 6,500 and 4,500 years ago, two closely related Neolithic cultures coexisted near the mouth of the Yangtze River; the Ma-chia-pang culture was located north of Hang-chou Bay, and the Ho-mu-tu culture was located to its south. This was a “subtropical, fresh-water wetlands environment of the coastal plain” that was “crisscrossed by rivers and streams and dotted with large and small lakes, ponds, and freshwater marshes” (Smith 1995: 124).

Large quantities of refuse reveal that these people were rice cultivators who probably kept domesticated pigs, dogs, and buffalo. Although the buffalo bones found there are no easier to distinguish as domesticated than at other sites, the unusually large number of bones does indicate a heavily exploited resource. Buffalo shoulder blades were lashed to wooden handles and used as spades, which were the main cultivation tools; no plows were used. Since these people evidently were skilled hunters and fishermen, it could be reasoned that large numbers of wild buffalo existed in the area and were intensively hunted.

If, however, the buffalo was domesticated, which seems to be the unanimous opinion of the excavators and analysts, it undoubtedly was kept for its meat (and shoulder blades); other uses are impossible to determine at this time. It is possible that even without the plow the buffalo might have been used to prepare the rice fields. As is practiced by many rice farmers in South China and Southeast Asia today, the soil can be

thoroughly puddled merely by driving one or more buffalo around in the flooded field (Ho 1975: 72).²¹ However, the finding of buffalo in association with cultivated rice does not necessarily imply that the animals were actually used in the cultivation process.

Radiocarbon dates place these buffalo in the fifth millennium B.C. Thus, if domesticated, they are the earliest ones known anywhere in the world (Pearson 1983; Chang 1986: 196–201, 208–11; Zhao and Wu 1986–7; Smith 1995: 124–7).²² Domesticated rice and buffalo have also been found as part of the Liang-chu culture (4000–2000 B.C.), which, although known since the 1930s, is now considered an outgrowth of the Ma-chia-pang culture in the same Yangtze Delta area.

The next appearance of the buffalo in a Neolithic context is in northern China at sites dated from the third millennium B.C. The earlier Neolithic cultures of North China (between 5500 and 2500 B.C.) appear to have been based on the hoe cultivation of millets and the keeping of domesticated pigs, dogs, and chickens. Although a few bovine bones have been found, their identity and domesticated status are uncertain (Ho 1975: 91–5; Smith 1995: 133–40). Bovine remains, claimed as domesticated, are more frequently encountered at later sites, with most identified as *Bos* sp. and only a few as *Bubalus* sp. (from Shaanxi Province).

By early historic times in northern China (second millennium B.C.), the domesticated buffalo seems to have been well established, as the remains from the Shang Dynasty sites near An-yang suggest. Bones identified as belonging to the species *Bubalus mephistopheles* were found in great abundance. Authorities believe that buffalo were able to exist this far north because the climate, as surmised from paleontological and palynological evidence and oracle bone inscrip-

tions, was warmer and wetter than it is today. There seems to be little doubt that these buffalo were fully domesticated, along with cattle, pigs, dogs, sheep, and goats.

Southeast Asia. Archaeology has not been able to offer much support for the rather popular assertion that peninsular Southeast Asia was the place of origin for the domestication of rice and, by association, the buffalo. The oldest presumably domesticated rice in Southeast Asia (from the Non Nok Tha and Ban Chiang sites in Thailand) dates only from the fourth millennium B.C., which makes it at least 1,000 (and possibly 3,000) years younger than the earliest Chinese rice. Demonstration of a great antiquity for the domesticated buffalo has also eluded archaeologists.

One early site in northern Vietnam has yielded buffalo remains dating from about 8000 B.C. Vietnamese archaeologists have suggested that this site, Bo-lum, and the many others belonging to the Bacsonian culture could be associated with the beginnings of agriculture (fruit trees, tubers, leguminous plants, and hill rice) and animal husbandry (dog and buffalo). But this claim appears to be based solely on the assumption that polished stone tools and pottery, which are found at the Bacsonian sites, imply an early Neolithic agricultural society. It is with greater confidence that we can approach the assertion that wet rice cultivation with buffalo-drawn plows existed in northern Vietnam by the time the Bronze/Iron Age Dong Son culture appears (600–400 B.C.). This claim is based on the finding of rice remains and bronze plowshares (Bellwood 1985: 273–5).

The most significant Southeast Asian evidence relating to the buffalo comes from the site of Ban Chiang in the Sakon Nakhon Basin of northeastern Thailand. An analysis of the faunal remains by Charles Higham and Amphan Kijngam (1979, 1985) has shown that the buffalo made its appearance here about 1600 B.C., coinciding with the first appearance of iron metallurgy. Remains of rice have been found throughout the site's occupation period, going back to about 3500 B.C.

Based on the size of the buffalo bones, which are considerably smaller than those of wild buffalo, and on the fact that the buffalo phalanges showed stress areas (indicative of plowing) similar to those found on modern work buffalo, the authors concluded not only that the Ban Chiang buffalo was domesticated but that it was used for plowing. No prehistoric plows have been found at this or any other site in Thailand. Higham and Kijngam reasoned that before 1600 B.C. the people of Ban Chiang were probably collecting wild rice and practicing swidden rice cultivation. With the introduction of the buffalo and iron, the agricultural system changed to an inundation system of wet rice farming. Two other sites in the Sakon Nakhon Basin show similar patterns, with buffalo absent from the earlier levels and appearing only in the later periods, apparently

about the same time (during the second millennium B.C.) as they do in Ban Chiang.

Ethnographic Evidence

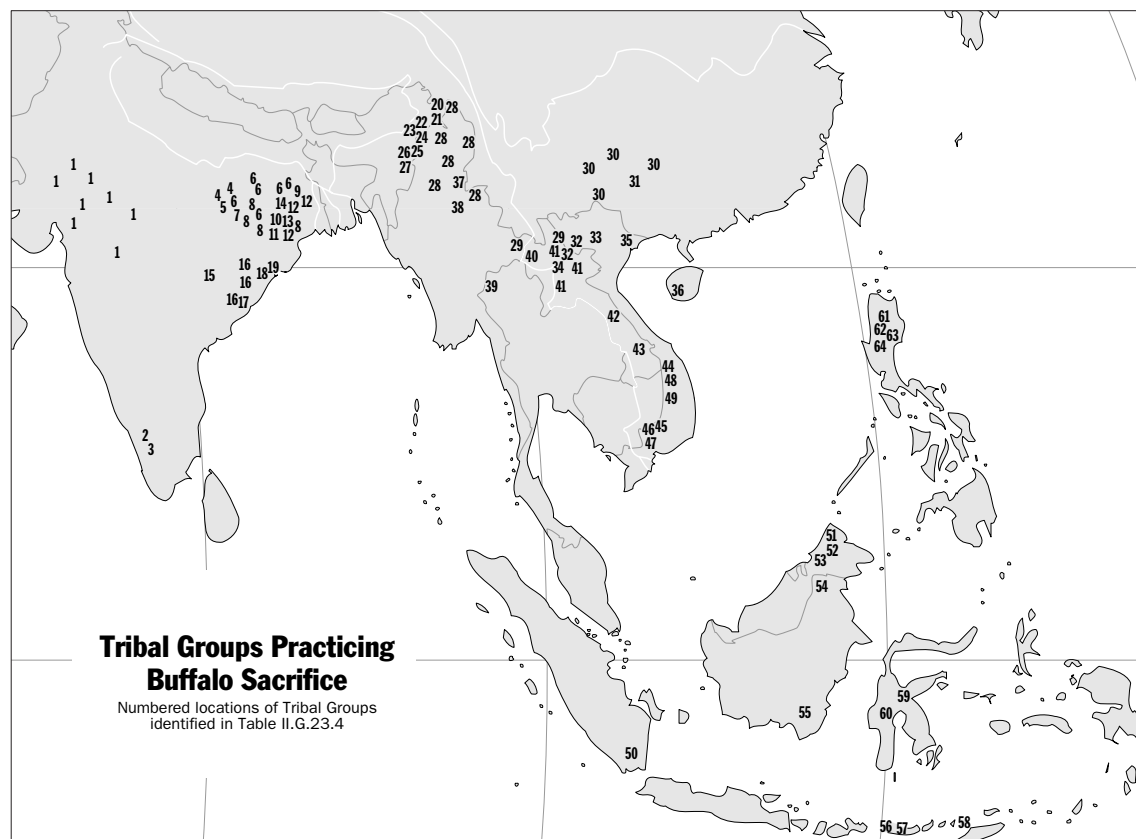
When archaeological findings are unsatisfyingly fragmentary, as is certainly the case with the buffalo, we search for any other possible indicators of the prehistoric past. The use of ethnographic evidence is based on the shaky assumption that primitive people living today or in the recent past can provide us with a glimpse, via extrapolation, of ways of life that may have existed long ago.²³ Drawing conclusions from such evidence must be done with great caution. Sometimes several different, yet equally reasonable, scenarios can be postulated.

A cursory investigation of the position of the buffalo among the primitive tribal societies of southern and eastern Asia reveals a surprisingly widespread pattern of use that might suggest an early motive for keeping and domesticating the buffalo. Most of these tribes inhabit hilly regions, practice the plowless shifting cultivation of dry rice, and do not drink milk.²⁴ Consequently, they usually do not keep domesticated animals for milk production, plowing, or other work.

Yet in spite of the absence of any of these economic uses, many groups do keep domesticated bovines. These animals are slaughtered and their meat eaten, but the motive is not simply the secular dietary desire for meat, which is met by the hunting of wildlife in the nearby forests. The major purpose for keeping these animals is for sacrifice, and their meat is eaten only at ceremonial feasts. Within the southern and eastern Asian realm, the buffalo is the bovine most frequently used for this purpose. Only a few tribal groups prefer other domesticated bovines as sacrificial victims, such as the zebu among a few tribes in India and the mithan among the groups inhabiting the eastern Himalayas and the India–Burma border region. Tribes practicing buffalo sacrifice are scattered throughout the Indian subcontinent, across mainland Southeast Asia, in parts of southern China, and in the hilly interiors of the islands of the Philippines (Map II.G.23.5 and Table II.G.23.4). All of the occasions for buffalo sacrifices are broadly related, in one way or another, to the desire to enhance or maintain the well-being of an individual or community – fertility in its broadest sense.

Economic or Religious Motivation

Given the obviously useful contributions of domesticated bovines to the economic life of humans, it is not surprising that most attempts to explain their origin seek rational and practical motives. Eduard Hahn (1896) was one of the earliest scholars to propose a thesis on cattle domestication that did not conform to the principles of materialistic rationalism. He was the first in a line of speculators who argued that cattle were originally domesticated not for milk, meat, or traction, but to provide sacrificial victims for the worshiped deities.



Map II.G.23.5. Tribal groups practicing buffalo sacrifice.

Table II.G.23.4. *Tribal groups practicing buffalo sacrifice (identification of numbers on Map II.G.23.5)*

Tribal name	Language affiliation	Source
India		
1. Bhil	Indo-Aryan (possible Munda origins)	Koppers and Jungblut 1942-5: 647-50; Naik 1956: 194-5; Nath 1960: 49, 209-10
2. Toda	Dravidian	Rivers 1906: 274, 287, 290, 354; Miles 1937: 94; India, Office of the Registrar General 1965: 77
3. Kota	Dravidian	Thurston 1909, 1: 85, 4: 25, 26
4. Korwa	Austroasiatic (Munda)	Dalton 1872: 228; Risley 1892, 1: 513; Majumdar 1944: 33, Das, Chowdhury, and Raha 1966: 130
5. Rautia	Austroasiatic (Munda)	Risley and Gait 1903: 409, 417
6. Oraon	Dravidian	Dalton 1872: 248; Risley 1892, 2: 145; Roy 1915: 191, 1928: 45-56, 68-9; Thakkar 1950: 37
7. Kharwar	Austroasiatic (Munda)	Risley and Gait 1903: 409, 417
8. Munda	Austroasiatic (Munda)	Tickell 1840: 700; Dalton 1872: 186; Risley 1892, 27: 103; Risley and Gait 1903: 144; Bradley-Birt 1910: 34; Crooke 1955: 16
9. Maler	Dravidian	Vidyarthi 1963: vii, 69, 139, 159, 160
10. Kharria	Austroasiatic (Munda)	Dalton 1872: 157; Risley 1892, 1: 468-9; Roy and Roy 1937, 2: 373-6
11. Juang	Austroasiatic (Munda)	Risley 1892, 1:353
12. Santal	Austroasiatic (Munda)	Dalton 1872: 212
13. Ho	Austroasiatic (Munda)	Tickell 1840: 799, 800; Dalton 1872: 186; Risley and Gait 1903: 144; Bradley-Birt 1910: 34; Majumdar 1927: 39, 1950: 256; Crooke 1955: 16
14. Bhumij	Austroasiatic (Munda)	Risley 1892, 1: 124
15. Muria Gond	Dravidian	Grigson 1949: 162-3; Elwin 1947: 66, 197, 657
16. Kond	Dravidian	Risley 1892, 1: 403; Risley and Gait 1903: 408; Imperial Gazetteer of India, 1908: 132; Thurston 1912: 201, 206; Russell 1916, 3: 473-9; Miles 1937: 82, 84, 87; Elwin 1950: 176; Das 1958, 27-30; Bailey 1960: 51

(continued)

Table II.G.23.4. (Continued)

Tribal name	Language affiliation	Source
India		
17. Gadaba	Austroasiatic (Munda)	Thurston 1909, 2: 250; Ramadas 1931: 172-3; Elwin 1954: xxv
18. Savara	Austroasiatic (Munda)	Dalton 1872: 147; Thurston 1912: 262-3; Sahu 1942: 82, 84; Bhat-tacharyya 1952: 9-15; Elwin 1955: 189-91, 229, 267-82, 284-6, 304, 305, 358-75, 378-86
19. Kuttia Kond	Dravidian	Elwin 1954: 352-3; Niggemeyer 1964: 184-211
20. Idu Mishmi	Sino-Tibetan (Tibeto-Burman)	Dalton 1872: 20; Baruah 1960: 22, 37, 80, 87
21. Singpho (Kachin)	Sino-Tibetan (Tibeto-Burman)	Dalton 1872: 16, 17
22. Tangsa	Sino-Tibetan (Tibeto-Burman)	Dutta 1959: 11, 30, 39, 59, 76-9
23. Konyak Naga	Sino-Tibetan (Tibeto-Burman)	Fürer-Haimendorf 1969: 55, 59
24. Chang Naga	Sino-Tibetan (Tibeto-Burman)	Dalton 1872: 16-17; Fürer-Haimendorf 1938: 206, 1946: 85, 127, 174, 176; Hutton 1965: 29, 34, 41
25. Kalyo-Kengyu Naga	Sino-Tibetan (Tibeto-Burman)	Dalton 1872: 16-17; Fürer-Haimendorf 1938: 209, 1946: 85, 127, 174, 176; Hutton 1965: 29, 34, 41
26. Rengma Naga (Eastern)	Sino-Tibetan (Tibeto-Burman)	Mills 1937: 4-11, 91-2
27. Tangkhul Naga	Sino-Tibetan (Tibeto-Burman)	Hodson 1911: 140, 149, 179
Mainland Southeast Asia and South China		
28. Kachin	Sino-Tibetan (Tibeto-Burman)	Enriquez 1923: 29-30; Leach 1964: 118-19, 143-7, 172-3; LeBar, Hickey, and Musgrave 1964: 13, 14, 17
29. Akha	Sino-Tibetan (Tibeto-Burman)	Scott and Hardiman 1900: 593; LeBar et al. 1964: 37
30. Miao of S. China	Miao-Yao	Mickey 1947: 78-80; Beauclair 1956: 29-31; LeBar et al. 1964: 71, 72
31. Chung-Chia	Thai	LeBar et al. 1964: 230
32. Black Thai	Thai	Halpern 1964: 29, 61, Table 19; LeBar et al. 1964: 221, 223
33. White Thai	Thai	Lévy 1959: 164; LeBar et al. 1964: 225
34. Laotian Thai	Thai	Archaimbault 1959: 156-61; Lévy 1959: 162-3; Halpern 1960: 64, 1964: 58, 61, 151, Table 19
35. Tho	Thai	LeBar et al. 1964: 233
36. Li	Kadai	LeBar et al. 1964: 241, 243
37. Wa	Austroasiatic (Mon-Khmer)	Scott 1896: 140, 145, 147; Scott and Hardiman 1900: 497, 505-6, 515; LeBar et al. 1964: 131
38. Palaung (Humai subgroup)	Austroasiatic (Mon-Khmer)	Cameron 1912: xxvii, xxxii, xxxvii
39. Lawa	Austroasiatic (Mon-Khmer)	Young 1962: 53, 54, 55; Kunstadter 1966: 132, 140
40. Lamet	Austroasiatic (Mon-Khmer)	Izikowitz 1951: 66, 67, 100-1, 106, 116, 118, 200-1, 232, 253, 303-5, 309, 316, 322-32, 357-8
41. Khmu	Austroasiatic (Mon-Khmer)	Halpern 1964: 38-9, 64; LeBar et al. 1964: 113-15
42. So	Austroasiatic (Mon-Khmer)	LeBar et al. 1964: 150
43. Alak	Austroasiatic (Mon-Khmer)	LeBar et al. 1964: 135
44. Sedang	Austroasiatic (Mon-Khmer)	Devereux 1937: 4-6; LeBar et al. 1964: 147
45. Mnong	Austroasiatic (Mon-Khmer)	LeBar et al. 1964: 156; Condominas 1966: 10, 11, 13, 1977: 19-20, 25-6, 58-80, 124, 203, 138-41, 1980: 219, 223, 228-40
46. Stieng	Austroasiatic (Mon-Khmer)	Gerber 1951: 230; LeBar et al. 1964: 158
47. Ma	Austroasiatic (Mon-Khmer)	LeBar et al. 1964: 153
48. Djarai	Austronesian (Cham)	Bertrand 1959: 56-7, 62-4, 113, 146, 182
49. Rhadé	Austronesian (Cham)	Bertrand 1959: 56-7, 62-4, 113, 146, 182; LeBar et al. 1964: 254-5
Island Southeast Asia		
50. Abung	Austronesian (Indonesian)	LeBar 1972: 37
51. Dusun	Austronesian (Indonesian)	Rutter 1922: 298; Evans 1923: 40; LeBar 1972: 149
52. Rungus Dusun	Austronesian (Indonesian)	LeBar 1972: 151
53. Idahan Merut	Austronesian (Indonesian)	LeBar 1972: 156
54. Kelabitic Merut	Austronesian (Indonesian)	LeBar 1972: 160, 163
55. Maanyan Dayak	Austronesian (Indonesian)	LeBar 1972: 192
56. Manggarai	Austronesian (Indonesian)	LeBar 1972: 81, 83
57. Endenese	Austronesian (Indonesian)	LeBar 1972: 87, 88
58. Alorese	Austronesian (Indonesian)	LeBar 1972: 97
59. Eastern Toradja	Austronesian (Indonesian)	Downs 1956: 77-91, 100-4; Geertz 1963: 70-6; LeBar 1972: 133
60. Southern Toradja	Austronesian (Indonesian)	Kennedy 1953: 42-3, 128, 136-48; LeBar 1972: 134, 136; Crystal 1974: 125-30, 132-6; Volkman 1980: 40-1, 111-17, 131-55; Nooy-Palm 1986: 5, 10-50, 62-76, 101-3, 116-17, 169-304
61. Kalinga	Austronesian (Indonesian)	Lawless 1975: 191-203; LeBar 1975: 93, 95
62. Bontoc	Austronesian (Indonesian)	Jenks 1905: 76; LeBar 1975: 97
63. Ifugao	Austronesian (Indonesian)	Barton 1946: 100, 194
64. Ibaloi	Austronesian (Indonesian)	Moss 1920: 286, 305-6

Those who followed with sympathetic support, similar premises, and further research included, most notably, Carl Sauer (1952: 88-94), Erich Isaac (1962), and Frederick and Elizabeth Simoons (1968: 234-58). These scholars have pointed out the widespread Eurasian distribution and antiquity of a complex of traits that link bovines, sacrificial practices, and fertility rites. Simoons and Simoons (1968: 261-2) postulated that taurine cattle were the first to be domesticated in this ceremonial context, followed by the buffalo in imitation of cattle, and then by the mithan in imitation of the buffalo.

In the case of the domestication of the water buffalo, we have very little concrete evidence and can offer only speculations that seem to accommodate the fragmented information we have. Nature has provided a massively built and formidable bovine that is ideally suited to tropical and swampy lowland environments. Paleontological evidence places the ancestor of this creature in southern and eastern Asia. Wild buffalo have seemingly adapted to a wide range of habitats as their distribution has expanded and contracted throughout the course of the Pleistocene and Holocene. Until the next discovery, currently available archaeological evidence points to southern China during the fifth millennium B.C. as the most probable place and time for the earliest domestication of the buffalo.

With the invention of the plow, the buffalo's contribution to the wet rice cultivation process was fully realized. Thus far, the earliest indication of the buffalo-plow-wet rice complex comes from northern Thailand about 3,600 years ago at the Ban Chiang site. It is postulated that this complex diffused into the Indo-Malaysian archipelago roughly 1,000 years later (Bellwood 1985: 205, 241-4). This line of thinking, which focuses on the evolving economic benefits accrued from the domesticated buffalo, would fulfill the requirements of materialistic rationalism. If ethnography can be transposed to the Neolithic, an alternative scenario can also be envisioned. Buffalo might have been initially tended for their religious and social value, rather than their economic value. A belief that the sacrificing of buffalo would be repaid with fertility and well-being could have motivated people to begin domestication. To conform with the archaeological evidence, it could be hypothesized that this buffalo sacrificing fertility cult initially developed in South China. Buffalo sacrificing has been reported among the early Thai (Eberhard 1968: 183-93, 216) and the traditional Miao peoples (Mickey 1947: 78-80; Beauclair 1956: 29-31) of that area. This sacrificial complex could have diffused southward into mainland Southeast Asia and from there westward into India.

By historic times, both buffalo systems - the primitive hill farming system with the sacrificial buffalo and the sophisticated lowland rice paddy system with the plow buffalo - were well established and widely distributed in southern and eastern Asia. Archaeology has not, as yet, been able to sort out the chronological

relationships between the two systems. The suggestion that the sacrificial buffalo system is antecedent to the plow buffalo system is based primarily on ethno-historical traditions that commonly identify the hill peoples of southern and eastern Asia as being the indigenous inhabitants and indicate that settlement and cultivation of the lowland plains came later. Without evidence to the contrary, we further assume that the recent cultural practices of these tribal peoples represent survivals from a much earlier time that have been preserved by the groups' relative isolation. Obviously, caution must accompany our speculation.

The popular archetypal image that many Westerners have of the water buffalo is as the lumbering beast of burden patiently pulling a plow and a frail Asian farmer through a flooded rice paddy. Although this may be an accurate depiction of the animal's economic value to humans, it may not reflect the original value conferred upon the water buffalo when it was first brought under domestication. We must consider an additional image of the buffalo as a powerful, virile symbol of fertility that is tied to a sacrificial post and about to be ceremonially slaughtered as an offering to the supernatural.

Biological, archaeological, and ethnographic evidence provides us with only a few fragments with which to construct images. As we try to fit the pieces together, we hope that the images that emerge, as preliminary and speculative as they may be, will be the result of objective, multidimensional analysis that has considered economic, religious, and all other relevant factors.

Robert Hoffpauir

Notes

1. Compilations of available knowledge of the buffalo have appeared from time to time, including Roderick MacGregor (1941), Bradford Knapp (1957), D. C. Rife (1959), W. Ross Cockrill (1967, 1974b, 1976), M. R. de Guzman and A. V. Allo (1975), and National Research Council (1981).
2. The present author's past work on the buffalo has included some cultural and historical considerations in an attempt to understand the current buffalo situation in South Asia (Hoffpauir 1968, 1974, 1977, 1982). Whereas several studies could be cited as examples of the cultural-historical approach to the analysis of domesticated animals, Frederick and Elizabeth Simoons's (1968) masterful study of the mithan (*Bos frontalis*) stands as the model of this approach.
3. Two earlier attempts to address specifically the domestication of the buffalo have revealed how little we know. Frederick Zeuner's often-cited survey, *A History of Domesticated Animals*, devoted only an 8-page portion of one chapter to the buffalo (1963: 245-52), and more recently, Cockrill (1984) contributed a short chapter on the buffalo to Ian L. Mason's anthology, *Evolution of Domesticated Animals*.
4. The term "bovine" refers to those animals that have been classified in the order Artiodactyla, suborder Ruminantia, family Bovidae, subfamily Bovidae. This includes European

- domestic cattle,* zebu cattle,* gaur, mithan,* banteng, Bali cattle,* *kouprey*, domestic yak,* wild yak, American bison, European bison, wild Asian buffalo, domesticated water buffalo,* *anoa*, *tamarau*, and African buffalo. (Asterisks [*] indicate the domesticated forms.)
5. The *anoa* (confined to the island of Sulawesi in Indonesia) and *tamarau* (confined to the island of Mindoro in the Philippines) are still usually classified under the *Bubalus* genus (*Bubalus depressicornis* and *Bubalus mindorensis*, respectively). The mountain *anoa* is sometimes given separate species status under the name *Bubalus quarlesi*. The African buffalo is now regarded as a more distantly related animal requiring a separate genus designation, *Syncerus caffer*.
 6. The literature on the environmental physiology of the buffalo is reviewed by Knapp (1957: 37–42) and Mason (1974).
 7. This section of the chapter focuses only on the present-day economic contributions of the buffalo. The importance of the buffalo in the socioreligious realm is of great cultural-historical significance, at least in this author's opinion, and is discussed later in the chapter.
 8. For comments on those rare situations of milking and milk consumption in this region, the reader is referred to Simoons (1970: 570–7) and Paul Wheatley (1965).
 9. Recent experimental introductions of very small numbers of buffalo have occurred in other parts of Africa – specifically in Mozambique, Tunisia, Madagascar, Uganda, Tanzania, Nigeria, Zaire, and Congo – but their current status is unknown (Porter 1991: 300).
 10. Esteem for the milk buffalo as an economic resource probably reaches its highest cultural expression in Southwest Asia among the Marsh Arabs (the Ma'dan) of southern Iraq. These "buffalo people" have developed an elaborate cultural system focused on the husbandry of buffalo. The reader is referred to the ethnographic descriptions by Wilfred Thesiger (1954, 1964).
 11. For a discussion of the prominent role played by wild buffalo in the cave art of Mesolithic India, see Yashodhar Mathpal (1984, 1985) and V. N. Misra (1985: 119–20).
 12. The often-cited identification of two teeth of Upper Pleistocene or early Holocene date found at Niah Cave in northern Borneo as belonging to *Bubalus* (Koenigswald 1958: 624–5) has been disputed by Lord Medway (1965: 162), who believes the teeth could just as well belong to *Bos javanicus*, the Banteng.
 13. It was once commonly thought that buffalo named *Bubalus antiquus* roamed over North Africa from the late Pleistocene to the Holocene, as revealed by bone finds and rock paintings and engravings created by Paleolithic and Neolithic hunters (Wulsin 1941). Current opinion, however, identifies these animals as members of the *Syncerus* genus (Zeuner 1963: 245–6). More than 100 years ago, Richard Lydekker (1898: 116) suggested that these buffalo might better be regarded as ancestral to the African buffalo.
 14. Unless otherwise mentioned, the sources used for the following discussion of archaeological sites are listed in Table II.G.23.3.
 15. For a description of the buffalo remains found at Harappa, see Bains Prashad (1936: 43–6).
 16. From Mohenjo-daro, Sir John Marshall (1931, 2: 386) listed only three seals showing buffalo, and E. J. H. Mackay (1937–8, 1: 298, 311, 330, 358–9) recorded one pottery figure, one bronze figure, eight seals, and two prism-amulets. From Harappa, Madho Sarup Vats (1940: 323) published only two fragmented seals.
 17. See, for example, Marshall (1931, 1: v), Prashad (1936: 8–9), M. S. Randhawa (1946: 10), Stuart Piggott (1950: 155–6), V. Gordon Childe (1957: 176), Kaj Birket-Smith (1965: 151), C. D. Darlington (1970: 55), Henry Hodges (1970: 252), and Cockrill (1984: 55).
 18. H. F. Friedrichs (as cited in Mackay 1937–8, 1: 670–1) argued that the presence of the trough before the buffalo, tiger, and rhinoceros indicates that these animals were kept under restraint, that is, that they were not as yet fully domesticated. The symbol was not needed in front of the zebu, sheep, goat, and elephant because these animals were more domesticated and possibly kept in pastures where a feeding trough would be unnecessary.
 19. A few authors have indicated their uncertainty about the buffalo's domesticated status in the Indus Valley (Zeuner 1963: 249; Bökönyi 1974: 151; Clason 1977: 268). One writer, D. D. Kosambi (1965: 60), has even expressed the opinion that the buffalo was probably not domesticated in the Indus Valley.
 20. Those who have supported the idea of a domesticated buffalo in ancient Mesopotamia include Seton Lloyd (1947: 7), Berthold Klatt (1948: 34), Samuel Noah Kramer (1967: 88), H. Epstein (1971: 569), S. Bökönyi (1974: 151), and Cockrill (1974b: 516, 1984: 56). Zeuner (1963: 249–50), although finding the evidence strongly suggestive of "tame, if not domesticated" buffalo, does admit that the question "remains to be settled." At least one author (Kingdon 1990: 136) has expressed the opinion that Mesopotamia is the place of earliest domestication of the buffalo. Presumably, these authors believe that the buffalo was kept primarily as a meat source, as no evidence has been presented for the buffalo being used for milk or work. Dominique Collon (1987: 187) says that buffalo – she does not specify wild or domesticated – may have been imported from the Indus Valley for the royal zoos during the time of Sargon of Akkad. Epstein (1971: 569) bolsters his argument for the existence of the domesticated buffalo with an ancient Assyrian clay tablet (from about 2000 B.C.) showing the head of a buffalo with "horns curved in a manner that is occasionally encountered in domestic buffaloes but never in the wild beast." Lynn White, Jr. (1974: 202) counters this argument by pointing out that wild African buffalo (*Syncerus*) have such horns.
 21. For a long time, the thought has been that the animal-drawn plow was not part of the original technique of wet rice cultivation but was added later (Pelzer 1945: 14). It has been hypothesized that the original form of rice cultivation involved broadcasting the seed onto naturally swampy or seasonally flooded lowland areas. The archaeological findings in the Yangtze Delta appear to be consistent with this hypothesis. The swidden cultivation of dry rice and the irrigation cultivation of wet rice, it is suggested, developed later (Harris 1974: 142; Gorman 1977: 338–40). Initially, no animals may have been used. Buffalo involvement in rice cultivation might have started with the use of the animal for puddling; perhaps only later was it attached to a plow. Wolfram Eberhard (1968: 215) contends that among the Thai of southern China, the practice of plowing, which was adopted very late, was anteceded by letting buffalo trample the fields. R. D. Hill (1977: 85, 98, 133, 161, 177) documents this practice during the nineteenth century in Malaya and Siam. It seems to have been widespread among many different groups, such as the Rhadé (LeBar, Hickey, and Musgrave 1964: 252) and Maà' Huang (Condominas 1980: 246) in Vietnam and Cambodia, the Savunese in the Savu Islands of Indonesia (LeBar 1972: 77), the Lampong in south Sumatra (Loeb 1935: 266), the Kalabits in north Borneo (Hose and McDougall 1912: 97–8), and the Kalingas of northern Luzon (Lawless 1975: 82–3).
 22. The only evidence of rice cultivation earlier than the Yangtze Delta finds is from the eastern end of the Yangtze

Valley at the site of Peng-tou-shan in the Hupei Basin, dated at between 6400 and 5800 B.C. No evidence of domesticated animals has been uncovered from this site as yet (Smith 1995: 130–1). Rice cultivation cannot be documented before 3600 B.C. in Southeast Asia or before 2500 B.C. in India (Bellwood et al. 1992). Thus far, the oldest evidence of both rice and buffalo domestication comes from the Yangtze Valley of China.

23. The value of this ethnographic approach, when carefully executed, is demonstrated by Simoons and Simoons's (1968) study of the mithan. In the absence of archaeological evidence, a skillful use of ethnographic evidence allowed Simoons and Simoons to develop a very convincing hypothesis for the domestication of this bovine.
24. For the purposes of this discussion, I employ the convention of speaking in the "ethnographic present." Some of the sources used describe conditions as they existed during the late nineteenth and early twentieth centuries. Subsequent acculturation has undoubtedly obliterated many of the (presumably) traditional practices originally observed and described.

Bibliography

- Agrawal, D. P. 1971. *The copper Bronze Age in India*. New Delhi.
- Aigner, Jean S. 1974. Pleistocene archaeological remains from South China. *Asian Perspectives* 16: 16–38.
1981. *Archaeological remains in Pleistocene China*. Munich.
- Allchin, Bridget, and Raymond Allchin. 1968. *The birth of Indian civilization: India and Pakistan before 500 B.C.* Baltimore, Md.
- Archaimbault, Charles. 1959. The sacrifice of the buffalo at Vat Ph'u (southern Laos). In *Kingdom of Laos*, ed. René de Berval, 156–61. Saigon.
- Badam, G. L. 1984. Pleistocene faunal succession of India. In *The evolution of the East Asian environment*, Vol. 2, ed. Robert Orr Whyte, 746–75. Hong Kong.
1985. The late Quaternary fauna of Inamgaon. In *Recent advances in Indo-Pacific prehistory*, ed. V. N. Misra and Peter Bellwood, 413–15. New Delhi.
- Bailey, F. G. 1960. *Tribe, caste, and nation: A study of political activity and political change in highland Orissa*. Manchester, England.
- Barton, R. F. 1946. The religion of the Ifugaos. *Memoir Series of the American Anthropological Association* 65: 1–219.
- Baruah, Tapan Kumar M. 1960. *The Idu Mishmis*. Shillong, India.
- Beauchair, Inez de. 1956. Culture traits of non-Chinese tribes in Kweichow Province, Southwest China. *Sinologica* 5: 20–35.
- Bellwood, Peter. 1985. *Prehistory of the Indo-Malaysian archipelago*. New York.
- Bellwood, P., R. Gillespie, G. B. Thompson, et al. 1992. New dates for prehistoric Asian rice. *Asian Perspectives* 31: 161–70.
- Bertrand, Gabrielle. 1959. *The jungle people: Men, beasts and legends of the Mōi country*. London.
- Bhattacharyya, Asutosh. 1952. Death rites, funeral ceremonies and idea of life after death among the hill Sora of Orissa. *Bulletin of the Department of Anthropology (Government of India)* 1: 1–16.
- Birket-Smith, Kaj. 1965. *The paths of culture*, trans. Karin Fennow. Madison, Wis.
- Bivar, A. D. H. 1969. *Catalogue of the western Asiatic seals in the British Museum*, Stamp seals, II: The Sassanian dynasty. London.
- Bliss, Frederick Jones. 1894. *A mound of many cities, or Tell el Hesi excavated*. London.
- Bodenheimer, F. S. 1960. *Animal and man in Bible lands*. Leiden, the Netherlands.
- Bökönyi, S. 1974. *History of domestic mammals in central and eastern Europe*. Budapest.
- Boriskovskii, P. I. 1970–1. Vietnam in primeval times, part VII. *Soviet Anthropology and Archeology* 9: 226–64.
- Bradley-Birt, F. B. 1910. *Chota Nagpore: A little-known province of the empire*. Second edition. London.
- Buffon, George-Louis Leclerc, Comte de. 1812. *Natural history, general and particular*, Vol. 7, trans. William Wood. London.
- Bühler, Georg, trans. 1879–82. *The sacred laws of the Aryas, as taught in the schools of Apastamba, Gautama, Vāsisbtha, and Baudhāyana*. Vol. 2:14 of *The sacred books of the East*, ed. F. Max Müller. Oxford.
1896. *The sacred laws of the Aryas, as taught in the schools of Apastamba, Gautama, Vāsisbtha, and Baudhāyana*. Apastamba and Gautama. Vol. 2:1 of *The sacred books of the East*, ed. F. Max Müller. Oxford.
- Cameron, A. A. 1912. A note on the Palaungs of the Kodaung hill tracts of the Momeik State. In *Census of India, 1911*, Vol. 9, Part 1, *Burma*, Appendix A, i–xlii. Delhi.
- Chand, Devi, trans. 1959. *The Yajur Veda*. Hoshiarpur, India.
- Chang, Kwang-chih. 1980. *S'ang civilization*. New Haven, Conn.
1986. *The archaeology of ancient China*. Fourth edition. New Haven, Conn.
- Childe, V. Gordon. 1957. *New light on the most ancient east*. New York.
- Clason, A. T. 1975. Archaeozoological study in India: Aspects of stock-breeding and hunting in prehistoric and early historic times. *The Eastern Anthropologist* 28: 1–12.
1977. Wild and domestic animals in prehistoric and early historic India. *The Eastern Anthropologist* 30: 241–394.
- Clutton-Brock, Juliet. 1981. *Domesticated animals from early times*. Austin, Tex.
- Cockrill, W. Ross. 1967. The water buffalo. *Scientific American* 217: 118–25.
- 1974a. Observations on skin colour and hair patterns. In *The husbandry and health of the domestic buffalo*, ed. W. Ross Cockrill, 48–56. Rome.
- ed. 1974b. *The husbandry and health of the domestic buffalo*. Rome.
1976. *The buffaloes of China*. Rome.
1984. Water buffalo. In *Evolution of domesticated animals*, ed. Ian L. Mason, 52–62. New York.
- Collon, Dominique. 1987. *First impressions: Cylinder seals in the ancient Near East*. Chicago.
- Condominas, George. 1966. The primitive life of Vietnam's mountain people. *Natural History* 75: 8–19.
1977. *We have eaten the forest: The story of a Montagnard village in the central highlands of Vietnam*, trans. Adrienne Foulke. New York.
1980. Agricultural ecology in the Southeast Asian savanna region: The Mnong Gar of Vietnam and their social space. In *Human ecology in savanna environments*, ed. David R. Harris, 209–51. London.
- Coon, Carleton S. 1955. *The story of man: From the first human*. New York.
- Creel, Herrlee Glessner. 1937. *The birth of China: A study of the formative period of Chinese civilization*. New York.
- Crooke, William. 1955. Dravidians (North India). In *Encyclopedia of religion and ethics*, Vol. 5, ed. James Hastings, 1–21. New York.

- Crystal, Eric. 1974. Cooking pot politics: A Toraja village study. *Indonesia* 18: 119-51.
- Dalton, Edward Tuite. 1872. *Descriptive ethnology of Bengal*. Calcutta.
- Darlington, C. D. 1970. The origins of agriculture. *Natural History* 79: 47-56.
- Das, Amal Kumar, Bidyut Kumar Roy Chowdhury, and Manis Kumar Raha. 1966. *Handbook on scheduled castes and scheduled tribes of West Bengal*. Calcutta.
- Das, Nityananda. 1958. Sukuli puja in a Kondh village. *Vanya-jati* 6: 27-30.
- Defen, Han, and Xu Chunhua. 1985. Pleistocene mammalian faunas of China. In *Palaeoanthropology and Palaeolithic archaeology in the People's Republic of China*, ed. Wu Rukang and John W. Olsen, 267-89. New York.
- Devereux, George. 1937. Functioning units in Ha(rh)ndea(ng) society. *Primitive Man* 10: 1-7.
- Downs, Richard Erskine. 1956. *The religion of the Bare'e-speaking Toraja of central Celebes*. The Hague.
- Duerst, J. Ulrich. 1908. Animal remains from the excavations at Anau. In *Explorations in Turkestan*, Vol. 2, ed. Raphael Pumpelly, 339-400. Washington, D.C.
- Dutta, Parul. 1959. *The Tangsas of the Namchik and Tirap Valleys*. Shillong, India.
- Eberhard, Wolfram. 1968. *The local cultures of South and East China*, trans. Alide Eberhard. Leiden, the Netherlands.
- El-Itriby, A. A. 1974. The buffalo in the Arab Republic of Egypt. In *The husbandry and bealth of the domestic buffalo*, ed. W. Ross Cockrill, 651-61. Rome.
- Elwin, Verrier. 1947. *The Muria and their ghotul*. London.
1950. *Bondo bigblander*. London.
1954. *Tribal myths of Orissa*. Bombay.
1955. *The religion of an Indian tribe*. London.
- Enriquez, C. M. 1923. *A Burmese Arcady*. London.
- Epstein, H. 1969. *Domestic animals of China*. Farnham Royal, England.
1971. *The origin of the domestic animals of Africa*, Vol. 1. New York.
1977. *Domestic animals of Nepal*. New York.
- Evans, Ivor H. N. 1923. *Studies in religion, folk-lore, and custom in British North Borneo and the Malay Peninsula*. Cambridge.
- Fairservis, Jan. 1971. *The roots of ancient India: The archaeology of early Indian civilization*. New York.
- FAO (Food and Agriculture Organization of the United Nations). 1992. *FAO production yearbook, 1991*, Vol. 45. Rome.
- Foote, R. Bruce. 1901. *Catalogue of the prehistoric antiquities*. Madras, India.
- Fox, Robert B. 1978. The Philippine Paleolithic. In *Early Paleolithic in South and East Asia*, ed. Fumiko Ikawa-Smith, 59-85. The Hague.
1979. The Philippines during the first millennium B.C. In *Early South East Asia: Essays in archaeology, history and historical geography*, ed. R. B. Smith and W. Watson, 227-41. New York.
- Frankfort, Henri. 1939. *Cylinder seals: A documentary essay on the art and religion of the ancient Near East*. London.
- Fürer-Haimendorf, Christoph von. 1938. Through the unexplored mountains of the Assam-Burma border. *The Geographical Journal* 91: 201-19.
1946. *The naked Nagas, head-hunters of Assam in peace and war*. Calcutta.
1969. *The Konyak Nagas: An Indian frontier tribe*. New York.
- Gade, D. W. 1970. Brazil's water buffalo. *Américas* 22: 35-9.
- Geertz, Hildred. 1963. Indonesian cultures and communities. In *Indonesia*, ed. Ruth T. McVey, 24-96. New Haven, Conn.
- George, J. C. 1955. Appendix II, Identification of bones. In *Report on the excavations at Nasik and Jorwe, 1950-51*, by H. D. Sankalia and S. B. Deo, 142-3. Poona, India.
- Gerber, Théophile. 1951. Coutumier Stien. *Bulletin de l'École Française d'Extrême-Orient* 45: 227-69.
- Gorman, Chester. 1971. The Hoabinhian and after: Subsistence patterns in Southeast Asia during the late Pleistocene and early recent periods. *World Archaeology* 2: 300-20.
1977. A priori models and Thai prehistory: A reconsideration of the beginnings of agriculture in southeastern Asia. In *Origins of agriculture*, ed. Charles A. Reed, 321-55. The Hague.
- Griffith, Ralph T. H., trans. 1963. *The hymns of the Rgveda*. 2 vols. Varanasi, India.
- Grigson, Wilfrid. 1949. *The Maria Gonds of Bastar*. London.
- Groves, C. P. 1981. Systematic relationships in the Bovini (Artiodactyla, Bovidae). *Zeitschrift für Zoologische Systematik und Evolutionsforschung* 19: 264-78.
- Guzman, M. R. de, and A. V. Allo. 1975. *The Asiatic water buffalo*. Taipei.
- Hahn, Eduard. 1896. *Die Haustiere und ihre Beziehungen zur Wirtschaft des Menschen*. Leipzig.
- Halpern, Joel M. 1960. Laos and her tribal problems. *Michigan Alumnus Quarterly Review* 67: 59-67.
1964. *Economy and society of Laos: A brief survey*. New Haven, Conn.
- Harris, David R. 1974. Rice and man in Southeast Asia. *Geographical Review* 64: 140-2.
- Higham, Charles F. W. 1977. Economic change in prehistoric Thailand. In *Origins of agriculture*, ed. Charles A. Reed, 385-412. The Hague.
- Higham, Charles F. W., and Amphan Kijngam. 1979. Ban Chiang and Northeast Thailand: The palaeoenvironment and economy. *Journal of Archaeological Science* 6: 211-33.
1985. New evidence for agriculture and stock-raising in monsoonal Southeast Asia. In *Recent advances in Indo-Pacific prehistory*, ed. V. N. Misra and Peter Bellwood, 419-23. New Delhi.
- Hill, R. D. 1977. *Rice in Malaya: A study in historical geography*. Kuala Lumpur, Malaysia.
- Hiltebeitel, Alf. 1978. The Indus Valley "Proto-Siva," reexamined through reflections on the goddess, the buffalo, and the symbolism of *vabanas*. *Anthropos* 73: 767-97.
- Ho, Ping-ti. 1975. *The cradle of the east: An inquiry into the indigenous origins of techniques and ideas of Neolithic and early historic China, 5000-1000 B.C.* Chicago.
- Hodges, Henry. 1970. *Technology in the ancient world*. New York.
- Hodson, T. C. 1911. *The Naga tribes of Manipur*. London.
- Hoffpauir, Robert. 1968. The domesticated buffalo in India and Pakistan: Its socio-economic distinctiveness as compared with the zebu. M.S. thesis, University of Wisconsin.
1974. India's other bovine: A cultural geography of the water buffalo. Ph.D. dissertation, University of Wisconsin.
1977. The Indian milk buffalo: A paradox of high performance and low reputation. *Asian Profile* 5: 111-34.
1978. Subsistence strategy and its ecological consequences in the Nepal Himalaya. *Anthropos* 73: 215-52.
1982. The water buffalo: India's other bovine. *Anthropos* 77: 216-38.
- Hose, Charles, and William McDougall. 1912. *The pagan tribes of Borneo*, Vol. 1. London.
- Hutton, John Henry. 1965. The mixed culture of the Naga tribe. *Journal of the Royal Anthropological Institute of Great Britain and Ireland* 95: 16-43.
- Imperial Gazetteer of India, Provincial series. 1908. *Central Provinces*. Calcutta.

- India. Office of the Registrar General. 1965. *Census of India, 1961*. Vol. 9, *Madras*. Part 5-C, Todas. Kelhi, India.
- Isaac, Erich. 1962. On the domestication of cattle. *Science* 137: 195-204.
- Izikowitz, Karl Gustav. 1951. *Lamet: Hill peasants in French Indochina*. Göteborg, Sweden.
- Jacob, Teuku. 1978. New finds of Lower and Middle Pleistocene Hominines from Indonesia and an examination of their antiquity. In *Early Paleolithic in South and East Asia*, ed. Fumiko Ikawa-Smith, 13-22. The Hague.
- Jenks, Albert Ernest. 1905. *The Bontoc Igorot*. Manila.
- Jia, Lanpo, and Weiwen Huang. 1985. The late Palaeolithic of China. In *Palaeoanthropology and Palaeolithic archaeology in the People's Republic of China*, ed. Wu Rukang and John W. Olsen, 211-23. New York.
- Jolly, Julius, trans. 1900. *The Institutes of Vishnu*. Vol. 7 of *The sacred books of the East*, ed. F. Max Müller. Oxford.
- Kennedy, Raymond. 1953. *Field note on Indonesia: South Celebes, 1949-50*. New Haven, Conn.
- Khan, M. Z. 1974. The buffaloes of Pakistan. In *The husbandry and health of the domestic buffalo*, ed. W. Ross Cockrill, 611-15. Rome.
- Kingdon, Jonathan. 1990. *Arabian mammals: A natural history*. New York.
- Kipling, John Lockwood. 1891. *Beast and man in India: A popular sketch of Indian animals in their relation with the people*. London.
- Klatt, Berthold. 1948. *Haustier und Mensch*. Hamburg.
- Knapp, Bradford. 1957. *A compilation of available data on the water buffalo*. Washington, D.C.
- Koenigswald, G. H. R. von. 1951. Introduction. In *Morphology of Solo man*, by Franz Weidenreich, 211-21. Anthropological papers of the American Museum of Natural History 43. New York.
1958. Remarks on the prehistoric fauna of the great cave at Niah. *The Sarawak Journal* 8: 620-6.
- Koppers, Wilhelm, and L. Jungblut. 1942-5. The water-buffalo and the zebu in central India. *Anthropos* 37-40: 647-66.
- Kosambi, D. D. 1965. *Ancient India: A history of its culture and civilization*. New York.
- Kramer, Samuel Noah. 1967. *Cradle of civilization*. New York.
- Krishnaswami, V. D. 1960. The Neolithic pattern of India. *Ancient India* 16: 25-64.
- Kunstadter, Peter. 1966. Living with Thailand's gentle Lua. *National Geographic* 130: 122-52.
- Kurtén, Björn. 1968. *Pleistocene mammals of Europe*. Chicago.
- Lal, B. B. 1954-5. Excavation at Hastinapura and other explorations in the upper Ganga and Sutlej basins 1950-52. *Ancient India* 10-11: 5-151.
- Lawless, Robert. 1975. The social ecology of the Kalingas of northern Luzon. Ph.D. dissertation, New School for Social Research.
- Leach, E. R. 1964. *Political systems of highland Burma: A study of Kachin social structure*. Reprint of 1954 edition. London.
- LeBar, Frank M. 1972. *Ethnic groups of insular Southeast Asia*. Vol. 1, *Indonesia, Andaman Islands, and Madagascar*. New Haven, Conn.
1975. *Ethnic groups of insular Southeast Asia*. Vol. 2, *Philippines and Formosa*. New Haven, Conn.
- LeBar, Frank M., Gerald C. Hickey, and John K. Musgrave. 1964. *Ethnic groups of mainland Southeast Asia*. New Haven, Conn.
- Lévy, Paul. 1959. The sacrifice of the buffalo and the forecast of the weather in Vientiane. In *Kingdom of Laos*, ed. René de Bervil, 162-73. Saigon.
- Lloyd, Seton. 1947. *Twin rivers: A brief history of Iraq from the earliest times to the present day*. Bombay.
- Loeb, Edwin M. 1935. *Sumatra: Its history and people*. Vienna.
- Lydekker, Richard. 1898. *Wild oxen, sheep, and goats of all lands, living and extinct*. London.
- MacGregor, Roderick. 1941. The domestic buffalo. *The Veterinary Record* 53: 443-50.
- Mackay, E. J. H. 1937-8. *Further excavations at Mobenjo-daro*. 2 vols. Delhi.
- Mahadevan, P. 1974. The buffaloes of Trinidad and Tobago. In *The husbandry and health of the domestic buffalo*, ed. W. Ross Cockrill, 698-704. Rome.
- Majumdar, D. N. 1927. Death and connected ceremonies amongst the Hos of Kolhan in Singbhum. *Journal and Proceedings of the Asiatic Society of Bengal* (new series) 23: 37-44.
1944. *The fortunes of primitive tribes*. Lucknow, India.
1950. *The affairs of a tribe: A study in tribal dynamics*. Lucknow, India.
- Marshall, John. 1931. *Mobenjo-daro and the Indus civilization*. 3 vols. London.
- Mason, Ian L. 1974. Environmental physiology. In *The husbandry and health of the domestic buffalo*, ed. W. Ross Cockrill, 88-104. Rome.
- Mathpal, Yashodhar. 1984. *Prehistoric rock paintings of Bhimbetka, central India*. New Delhi.
1985. The hunter-gatherer way of life depicted in the Mesolithic rock paintings of central India. In *Recent advances in Indo-Pacific prehistory*, ed. V. N. Misra and Peter Bellwood, 177-83. New Delhi.
- Meacham, William. 1977. Continuity and local evolution in the Neolithic of South China: A non-nuclear approach. *Current Anthropology* 18: 419-40.
1983. Origins and development of the Yüeh coastal Neolithic: A microcosm of culture change on the mainland of East Asia. In *The origins of Chinese civilization*, ed. David N. Keightley, 147-75. Berkeley, Calif.
- Medway, Lord. 1965. *Mammals of Borneo: Field keys and an annotated checklist*. Singapore.
1972. The Quaternary mammals of Malesia: A review. In *The Quaternary era in Malesia*, ed. P. Ashton and M. Ashton, 63-98. Transactions of the Second Aberdeen-Hull Symposium on Malesian Ecology, Aberdeen 1971. Department of Geography, University of Hull, Miscellaneous Series No. 13. Hull, England.
- Mickey, Margaret Portia. 1947. *The Courie Shell Miao of Kweichow*. Papers of the Peabody Museum of American Archaeology and Ethnology, Harvard University, 32. Cambridge, Mass.
- Miles, Arthur. 1937. *The land of the lingam*. London.
- Mills, James Philip. 1937. *The Rengma Nagas*. London.
- Misra, V. N. 1985. Microlithic industries in India. In *Recent advances in Indo-Pacific prehistory*, ed. V. N. Misra and Peter Bellwood, 111-22. New Delhi.
- Moss, C. R. 1920. Nabaloi law and ritual. *University of California Publications in American Archaeology and Ethnology* 15: 207-342.
- Movius, Hallam L., Jr. 1944. *Early man and Pleistocene stratigraphy in southern and eastern Asia*. Papers of the Peabody Museum of American Archaeology and Ethnology, Harvard University, 19. Cambridge, Mass.
- Murray, James A. H., Henry Bradley, W. A. Craigie, and C. T. Onions. 1933. *The Oxford English dictionary*, Vol. 1. Oxford.
- Naik, T. B. 1956. *The Bbils: A study*. Delhi.
- Nath, Y. V. S. 1960. *Bbils of Ratanmal: An analysis of the*

- social structure of a western Indian community. Baroda, India.
- National Research Council, Panel on Water Buffalo (U.S.). 1981. *The water buffalo: New prospects for an underutilized animal*. Washington, D.C.
- Niggemeyer, Hermann. 1964. *Kuttia Kond: Dschungel-Bauern in Orissa*. Munich.
- Nooy-Palm, Hetty. 1986. *The Sa'dan-Toraja: A study of their social life and religion*. Dordrecht, the Netherlands.
- Paul the Deacon. 1907. *History of the Langobards*, trans. William Dudley Foulke. Philadelphia, Pa.
- Pearson, Richard. 1983. The Ch'ing-lien-kang culture and the Chinese Neolithic. In *The origins of Chinese civilization*, ed. David N. Keightley, 119-45. Berkeley, Calif.
- Pelzer, Karl J. 1945. *Pioneer settlement in the Asiatic tropics: Studies in land utilization and agricultural colonization in southeastern Asia*. New York.
- Pfeffer, Pierre. 1968. *Asia: A natural history*. New York.
- Phillips, Ralph W., Ray G. Johnson, and Raymond T. Moyer. 1945. *The livestock of China*. Washington, D.C.
- Piggott, Stuart. 1950. *Prehistoric India to 1000 B.C.* Harmondsworth, England.
- Pliny. 1855. *The natural history of Pliny*, Vol. 2, trans. John Bostock, and H. T. Riley. London.
- Pokee, Sohn. 1984. The palaeoenvironment of Middle and Upper Pleistocene Korea. In *The evolution of the East Asian environment*, Vol. 2, ed. Robert Orr Whyte, 877-93. Hong Kong.
- Polikhronov, D. St. 1974. The buffaloes of Bulgaria. In *The husbandry and health of the domestic buffalo*, ed. W. Ross Cockrill, 709-30. Rome.
- Porter, Valerie. 1991. *Cattle: A handbook to the breeds of the world*. New York.
- Prashad, Bani. 1936. *Animal remains from Harappa*. Memoirs of the Archaeological Survey of India, No. 51. Delhi.
- Qiu, Zhonglang. 1985. The Middle Palaeolithic of China. In *Palaeoanthropology and Palaeolithic archaeology in the People's Republic of China*, ed. Wu Rukang and John W. Olsen, 187-210. New York.
- Ramadas, G. 1931. The Gadabas. *Man in India* 11: 160-73.
- Randhawa, M. S. 1946. Role of domesticated animals in Indian history. *Science and Culture* 12: 5-14.
- Rao, S. R. 1962-3. Excavation at Rangpur and other explorations in Gujarat. *Ancient India* 18-19: 5-207.
- Rea, Alexander. 1915. *Catalogue of the prehistoric antiquities from Adichanallur and Perumbair*. Madras, India.
- Rife, D. C. 1959. *The water buffalo of India and Pakistan*. Washington, D.C.
- Risley, H. H. 1892. *The tribes and castes of Bengal*. 2 vols. Calcutta.
- Risley, H. H., and E. A. Gait. 1903. *Census of India, 1901*. Vol. 1, *India. Part 1, Report*. Calcutta.
- Rivers, W. H. R. 1906. *The Todas*. London.
- Rouse, John E. 1970. *World cattle*. 2 vols. Norman, Okla.
- Roy, Sarat Chandra. 1915. *The Oraons of Chota Nagpur: Their history, economic life, and social organization*. Ranchi, India.
1928. *Oraon religion and customs*. Calcutta.
- Roy, Sarat Chandra, and Ramesh Chandra Roy. 1937. *The Kbarias*. 2 vols. Ranchi, India.
- Russell, R. V. 1916. *The tribes and castes of the central provinces of India*. 4 vols. London.
- Rutter, Owen. 1922. *British North Borneo: An account of its history, resources and native tribes*. London.
- Sahu, Lakshminarayana. 1942. *The bill tribes of Jeyapore*. Cuttack, India.
- Salerno, A. 1974. The buffaloes of Italy. In *The husbandry and health of the domestic buffalo*, ed. W. Ross Cockrill, 737-47. Rome.
- Sankalia, H. D. 1946. *Investigations into prehistoric archaeology of Gujarat*. Baroda, India.
1977. *Prehistory of India*. New Delhi.
1978. *Pre-historic art in India*. Durham, N.C.
- Sastri, K. N. 1957. *New light on the Indus civilization*, Vol. 1. Delhi.
- Sauer, Carl O. 1952. *Agricultural origins and dispersals*. New York.
- Scott, James George. 1896. The wild Wa: A head-hunting race. *Asiatic Quarterly Review* (third series) 1: 138-52.
- Scott, James George, and J. P. Hardiman. 1900. *Gazetteer of upper Burma and the Shan State*, Part 1, Vol. 1. Rangoon.
- Simoons, Frederick J. 1970. The traditional limits of milking and milk use in southern Asia. *Anthropos* 65: 547-93.
- Simoons, Frederick J., and Elizabeth S. Simoons. 1968. *A ceremonial ox of India: The mithan in nature, culture, and history*. Madison, Wis.
- Smith, Bruce D. 1995. *The emergence of agriculture*. New York.
- Soundara Rajan, K. V. 1958. Studies in the stone age of Nagarjunakonda and its neighbourhood. *Ancient India* 14: 49-113.
- Sowerby, Arthur de Carle. 1935. The domestic animals of China. *The China Journal* 23: 233-43.
- Srinivasan, Doris. 1975-6. The so-called Proto-Siva seal from Mohenjo-daro: An iconological assessment. *Archives of Asian Art* 29: 47-58.
- Sullivan, Herbert P. 1964. A re-examination of the religion of the Indus civilization. *History of Religions* 4: 115-25.
- Terra, H. de, and P. Teilhard de Chardin. 1936. Observations on the upper Siwalik formation and later Pleistocene deposits in India. *Proceedings of the American Philosophical Society* 76: 791-822.
- Thakkar, A. V. 1950. *Tribes of India*. Delhi.
- Thapar, B. K. 1957. Maski 1954: A Chalcolithic site of the southern Deccan. *Ancient India* 13: 4-142.
- Thesiger, Wilfred. 1954. The Ma'dan or marsh dwellers of southern Iraq. *Journal of the Royal Central Asian Society* 41: 4-25.
1964. *The Marsh Arabs*. London.
- Thurston, Edgar. 1909. *Castes and tribes of southern India*. 7 vols. Madras, India.
1912. *Omens and superstitions of southern India*. London.
- Tickell, Lieut. 1840. Memoir on the Hodesum (Improperly called Kolehan). *Journal of the Asiatic Society of Bengal* 9: 694-710.
- Vats, Madho Sarup. 1940. *Excavations at Harappa*, Vol. 1. Delhi.
- Vidyarthi, L. P. 1963. *The Maler: A study in nature-man-spirit complex of a bill tribe in Bihar*. Calcutta.
- Volkman, Toby Alice. 1980. The pig has eaten the vegetables: Ritual and change in Tana Toraja. Ph.D. dissertation, Cornell University.
- Ward, William Hayes. 1910. *The seal cylinders of western Asia*. Washington, D.C.
- Wetherill, Elkins. 1972. A preliminary report on faunal remains from Chasen. Appendix to Excavations at Chasen, Thailand, 1968 and 1969: A preliminary report, by Bennet Bronson and George F. Dales. *Asian Perspectives* 15: 44-5.
- Wheatley, Paul. 1965. A note on the extension of milking

practices into Southeast Asia during the first millennium A.D. *Anthropos* 60: 577-90.

Wheeler, Mortimer. 1959. *Early India and Pakistan, to Ashoka*. New York.

1968. *The Indus civilization*. Third edition. Cambridge.

White, Lynn, Jr. 1974. Indic elements in the iconography of Petrarch's "Trionfo della morte." *Speculum* 49: 201-21.

Whyte, R. O. 1968. *Land, livestock and human nutrition in India*. New York.

Whyte, R. O., and M. A. Mathur. 1974. The buffalo in India. In *The husbandry and health of the domestic buffalo*, ed. W. Ross Cockrill, 548-68. Rome.

Wu, Xinzhi, and Linghong Wang. 1985. Chronology in Chinese palaeoanthropology. In *Palaeoanthropology and palaeolithic archaeology in the People's Republic of China*, ed. Wu Rukang and John W. Olsen, 29-51. New York.

Wulsin, Frederick R. 1941. *The prehistoric archaeology of northwest Africa*. Papers of the Peabody Museum of American Archaeology and Ethnology, Harvard University, 19. Cambridge, Mass.

Young, Gordon. 1962. *The hill tribes of northern Thailand*. Bangkok.

Yule, Henry, and A. C. Burnell. 1903. *Hodson-Jobson: A glossary of colloquial Anglo-Indian words and phrases, and of kindred terms, etymological, historical, geographical and discursive*. London.

Zeuner, Frederick E. 1963. *A history of domesticated animals*. New York.

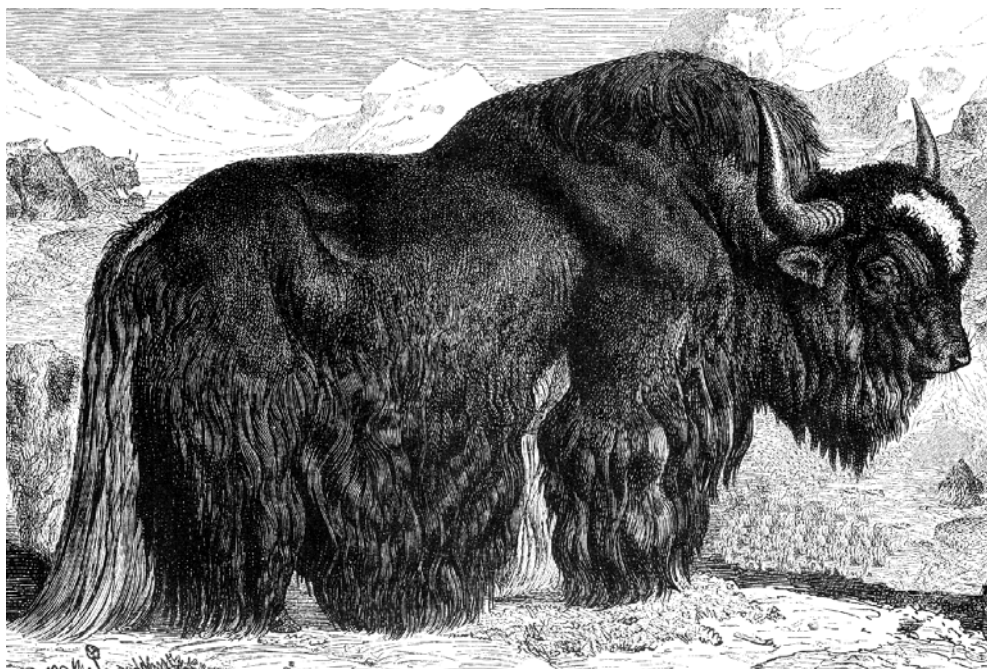
Zhao, Songqiao, and Wei-Tang Wu. 1986-7. Early Neolithic Hemodu culture along the Hangzhou estuary and the origin of domestic paddy rice in China. *Asian Perspectives* 27: 29-34.

II.G.24 🐃 Yak

Despite recent interest in domesticated animals in general and bovines in particular, there has been little systematic study of yak.¹ Nevertheless, enough is known to warrant a closer consideration of the role these cattle play in the culture, diet, and ecology of the several peoples who exploit them in some of the harshest and most difficult environments of Asia.

Taxonomy, Description, and Habitat

Yak are members of the subfamily of cattle Bovinae.² Although the genetic relationships among its members are not precisely understood, Herwart Bohlken (1958: 167-8, 1958-60: 113-202) argues that yak belong to the genus *Bos* and subgenus *Poëphagus*. Moreover, Bohlken draws a further distinction between wild yak, *Bos [Poëphagus] mutus* Przhelval'skii (1883), and domesticated yak, *Bos [Poëphagus] grunniens* Linnaeus (1766). Although wild and domesticated yak are interfertile, domesticated yak can be crossed with a variety of other cattle, including common cattle (*Bos taurus*), and zebu (*Bos indicus*) to produce hybrids of various types.³



Yak

Both wild and domesticated yak are massively built, with barrel-shaped bodies carried on legs that are quite short but solid (see Figure II.G.24.1). Although yak have no humps (as zebu do), they do have a dorsal ridgelike prominence that adds to their massive appearance. From this prominence, a short heavy neck slopes downward, ending in a large head with a broad, flat forehead, large eyes, and small ears.

Probably the most conspicuous features of the yak, however, are its horns and hair. The horns are large, dark, and double-curved. That is, they emerge horizontally from either side of the head, curve and extend first upward, then backward. The hair is coarse and shaggy and covers the animal's body almost entirely. The hair is abundant and is generally 4 or more inches in length. Fringing the chest, lower shoulders, sides, flanks, and thighs of the yak is even longer hair, which almost reaches the ground. The tail, too, is entirely covered by long hair, giving it a pronounced bushy appearance.

Wild yak are characteristically black or dark brown, occasionally with lighter colors on the forehead and along the back. Domesticated yak, in contrast, are much more variable in color. They range from black to white, and even piebalds are common.

Large numbers of wild yak once roamed over much of Tibet and the high Himalayas, though they have decreased in numbers and range. Today, only small herds occur, and only in certain restricted and inhospitable places, including small portions of the great mountain ranges of Inner Asia (the Tien Shan, Kunlun Shan, Pamirs, Karakorum) and the most remote areas of the Greater Himalayas. They are also found in small numbers in the dry Changtang, a vast, virtually uninhabited plateau that sweeps across northern Tibet in a great arc, extending from the Tibet-India border in the west to the Koko Nor region in the east.

In the better watered areas of the Changtang, small herds of wild yak give way to the large herds of domesticated yak managed by pastoralists. Domesticated yak also thrive in the more moist regions of Mongolia and central Asia and in the deep river valleys and small plains of southern and eastern Tibet, as well as along the high, south-facing flank of the Himalayas. Here, temperatures are higher, rainfall is much greater and more predictable, and vegetation is more abundant. These are regions occupied by farming folk, who are far more numerous than the pastoral nomads of the plateaus.

Yak in Culture and Diet

Wherever yak are kept, they are a central element in human ecology and adaptation. Farmers, for example, use these animals for a variety of agricultural purposes, including plowing and threshing. Among pastoral nomads, yak are a source of hair for tent-cloth and may be employed in caravan work. For

transhumant populations in mountainous regions, who combine certain aspects of pastoral and agricultural economies, yak are employed in these and other ways to wrest a living from often fragile environments. But most important, perhaps, to those who keep yak is the role the animal plays in the direct provision of food for human consumption and nutrition. In this regard, any detailed consideration of the wide variety of useful dietary products supplied by yak must focus on meat, milk, milk products, and blood.

Meat

Although most yak-keeping folk profess Buddhism, a religion that encourages a vegetarian diet by forbidding the slaughter of animals, few abstain from meat eating. Quite the opposite seems to be the case. Flesh foods are relished by most of these people, and every attempt is made to include at least some meat in the main meal each day. This is not always possible, however, because in most regions meat is a seasonal commodity and, in any case, is expensive.

The Buddhist prohibition of slaughter stems from a central concept in Buddha's teachings: *ahimsa*. Slaughter and meat eating compromise Buddhist yak-herders and are a source of guilt, for they contribute to the suffering and death of yak. To escape guilt, people slaughter yak in ways that rationalize and minimize responsibility. For example, calves may be denied the milk of their mothers. Since such animals starve and blood is not shed, death is considered a matter of fate and not the result of human action.

When slaughter by starvation is not practical, other methods are employed. Adult yak, for example, can be suffocated with a cord tied tightly around the animal's mouth and nose. And according to Marion H. Duncan (personal communication), Tibetans of Kham in eastern Tibet fit a leather sack tightly around the animal's muzzle, thereby cutting off air.

Perhaps the most common method of minimizing guilt is to give yak over to a hereditary group of professional butchers for slaughter. Such butchers absolve others of moral responsibility in the affair and are considered damned for their transgressions. Many butchers form a distinct caste in Buddhist societies. Others, such as those of Lhasa, are Moslems, originally from Ladakh or Kansu, who do not adhere to the *ahimsa* concept.

Although the slaughter of yak by butchers takes place throughout the year, especially in larger settlements, slaughtering by farmers and nomads is a seasonal affair. Most are killed in late autumn or early winter when the animals are well fleshed from spring and summer grazing. This is the time of year when winter temperatures facilitate preservation and storage of meat; the timing of the slaughter also reflects an attempt, in regions of limited pasturing opportunities, to minimize overgrazing and herd losses by careful control of herd size.



Figure II.G.24.1. Domesticated yak. (Drawing by Hildreth Burnett.)

Selection of yak to be slaughtered in late autumn or early winter is not random. Rather, each animal is carefully considered, and those deemed least able to survive the harsh winter months are culled out first. To these are added others that are considered poor economic investments. Such yak are usually the male cattle, often yak bulls too old, too sick, or too weak for traction or burden. Female yak are selected if they are dry or past the age of breeding. Selective slaughter, thus, not only provides an abundance of meat but also enhances the ability of the remaining cattle to survive the deprivations of winter. In addition, it results in a more efficient allocation of pasturing and other resources and a more efficient and economical herd.

Meat not eaten immediately after slaughter must be preserved or traded. If fresh meat is traded, distances involved are very limited – within a village, for example, to folk who have little access to meat and who pay cash, barter goods, or provide services in exchange. If properly preserved, however, such meat can be kept for months, even years, and traded over great distances.

Yak meat is preserved in a variety of ways. The lower temperatures of the slaughtering season may be used to freeze it. The meat is also salted and

smoked, but drying is by far the most common method of preservation. Carcasses of slaughtered yak may be simply cut into joints, which are hung outdoors. The joints slowly dry, often shrinking to half their original size.

More commonly, the beef is cut into thin strips, which are set out in the sun to dry, sometimes on cloths spread over the ground, more often on racks erected for this purpose. Beef jerked in this way is very popular in Tibet and the high Nepal Himalaya.

Not only does drying preserve meat, it also reduces its weight, allowing more to be carried on long journeys. In wintertime, dried meat is an important item of barter and sale. This is especially true for pastoral nomads, who use meat and other animal products to trade for the carbohydrate-rich plant foods of farmers and transhumant folk.

Meat also plays a prominent role in the diet of most yak-keeping peoples as a ceremonial or feast food. Among more prosperous families, meat of some kind is eaten nearly every day, but because of the expense involved, poorer families eat meat only occasionally. Whenever possible, however, even the poorest will try to offer a meat dish to visitors and guests. When unable to do so, apologies are given or excuses made.

Not everyone eats yak meat, however, and even those who do often abstain from certain kinds of flesh or from flesh prepared in proscribed ways. Some Tibetans, for example, consume the flesh of yak and female hybrids eagerly but refuse to eat the flesh of male hybrids. When asked why, they reply that male hybrids are sinful animals and their flesh is polluting.

Yak meat that has been roasted, broiled, or fried is considered polluting and impure by virtually all Tibetans and Himalayan Buddhists.⁴ Some Tibetan nomads believe that consumption of flesh prepared in these ways will result in sickness or bad luck – the punishment of gods angered by the odor of scorched flesh. Although many farmers and transhumant peoples agree with this pastoral nomadic belief, perhaps in an attempt to emulate this more prestigious group, others strike a compromise. According to them, the indoor roasting, broiling, and frying of meat produces odors that house gods find offensive. If, however, meat is prepared in these ways outside the home, it can be eaten with no harmful effects. In any case, the most common method of cooking yak and hybrid meat is boiling, a method acceptable to all. But flesh is also eaten uncooked, whether fresh or preserved.

In a few areas of Tibet bordering on India, Tibetans have been affected by Hindu views of the sacredness of common cattle. In some cases, Hindus themselves have transferred their views concerning common cattle to yak and hybrids, and in areas under their political influence such flesh has been (at one time or another), difficult or impossible to get. In other cases, Tibetans have taken over Hindu views. Some Tibetans of western Tibet, for example, were reported to view yak as sacred and avoided eating its flesh at all costs. Other Tibetans of the same region, however, who were less affected by Hindu ways, ate yak meat freely (Sherring 1906).

Some inhabitants of the Nepal Himalaya have been similarly influenced by Hindu ideas. The Gurung and Thakali are two such groups. The legends of the Gurungs, a tribal people of west central Nepal, indicate that they once consumed yak flesh freely, but according to Donald Messerschmidt (personal communication), present-day Hinduized Gurung consider this practice abhorrent because they have transferred Hindu ideas concerning common cattle to yak.

The Thakali of the Kali Gandaki Valley gave up eating the flesh of yak in the nineteenth century in an attempt to elevate the social status of their caste and to facilitate establishment of social relations with dominant members of Nepal's Hindu ruling class. Tradition maintains that Harkaman Subba Sherchan (1860–1905), a customs contractor, initiated the banning of yak flesh (Fürer-Haimendorf 1966: 144–5).

Milk

Because of the feelings of guilt associated with the consumption of yak flesh, meat is included in a category of “black” foods, considered unceremonious, harmful, or improper. Milk and milk products, how-

ever, are classed as “white” foods, suitable for all occasions and persons. Most yak herders consider milk, in both its fresh and processed states, a tasty food to be consumed eagerly by all ages. But while these views are associated with milk in general, distinctions are made among a variety of milk types, each the product of different mammals. Furthermore, each type of milk occupies a position in a hierarchy of milk preference that varies little, if at all, from place to place.

Among Tibetans and Himalayan folk, for example, there exists a strikingly consistent pattern or hierarchy of milk preference from place to place and group to group. Above all others, milk of yak is preferred, not only for its flavor and richness but also because it is considered especially healthful. Milk of hybrid cows is thought to be less tasty than yak milk but is preferred over that of common cattle (the other component of hybrid parentage).

Distinctions, however, are made by those more familiar with hybrids. The Sherpa of Nepal, for example, distinguish between milk of the two major hybrid types and prefer milk of hybrids (dimdzo) sired by common bulls to that of hybrids (urangdzo) sired by yak bulls. In any case, milk of common cows is not especially liked but is considered superior to the milk of backcross cows and ewes and is consumed when more highly prized milk is not available. Milk of female goats is least preferred and is seldom consumed by Tibetans and Himalayan folk, even the poorest.

This pattern of preference is generally explained in terms of the varying quality of milk types. Quality is defined as “thickness” or “richness.” Such terms refer to the fat content of the various milks, for analyses of these milks do indeed show substantial differences in fat content. This content may range widely, however, depending on such factors as the season when the milk was obtained, grazing conditions at the time of milking, and availability and utilization of supplementary feed. Thus, the fat content by volume of yak milk has been assayed by some to be between 4.8 percent and 16 percent (Mittaine 1962: 693; Schley 1967: 45–6, 78). But a study by Peter Schley of experiments on yak, hybrids, and common cattle in the Soviet Union has demonstrated that at any given time, fat content of yak milk is 18 percent to 28 percent greater than that of hybrid cow milk and 39 percent to 97 percent greater than the milk of common cows. Furthermore, the fat content of hybrid cow milk is, according to Schley's figures, 13 percent to 65 percent greater than that of common cows (Schley 1967: 78).

With respect to fat content, the obvious superiority of yak milk to that of hybrids and common cows is somewhat offset by the inability of yak to produce as great a volume of milk as hybrid and common cows. But, as with the fat content of milk, production figures for yak, hybrids, and common cows vary from place to place and season to season. Grazing conditions account for a goodly part of these variations, but milk production also varies with age of milch cows. Yak and hybrid

cows, for example, begin lactating at 2 or 3 years of age. As they grow older, and until they reach the age of 10, the lactation period lengthens from approximately 230 days to 300 days. Naturally, this lengthening is reflected in higher production of milk. After this maximum is reached, milk production declines.

Whatever the actual milk production figures are for yak and for hybrids and common cows kept under similar conditions, field observations in the Nepal Himalaya reveal several constants that should be noted. First, hybrid cows produce twice as much milk as yak and 50 percent more than common cows. The latter, however, consistently produce more than yak by about 30 percent.

The second constant reflects the season of milk production. Milk yields for yak, hybrids, and common cows increase as summer waxes and grazing improves, then begin to decline in late summer. In midwinter, lactation nearly ceases, though hybrids tend to be able to lactate during this period somewhat better than yak or common cows. Percentage of fat by volume, however, does not decline with the seasonal decline in milk quantity. Rather, it continues to climb, reaching its maximum when milk yields are lowest. These data confirm the findings of Schley and others but conflict with certain other sources.⁵

Tibetans and others do make some efforts to increase the milk production of their cattle. For example, they use stud that derive from lineages renowned for their milking characteristics. Another technique involves supplemental feeding of lactating yak and hybrids. A number of fodders are used for this purpose, including grasses and leaves collected in pasture and transported to the agricultural village or nomadic camp, where the feed is given to animals in the evening or stored for future use. Also used are fodder crops such as radishes and the straw of harvested grains. And some Tibetan nomads occasionally cultivate fields of planted and volunteer fodder crops for just such use.

Among the more interesting Tibetan customs employed to maximize milk yields is the feeding of salt to lactating cows; more frequent milking, especially as the grazing season progresses; restricting nursing by calves; and stimulating the milk let-down response in stubborn females.⁶ In this last, several methods are employed, among them use of calf substitutes, calf dolls, and vaginal insufflation.⁷

Milking of yak and hybrids is done almost exclusively by women and girls; only occasionally will men and boys involve themselves in this work. Whoever does it, the milking procedure seldom varies. The milker always squats beside the cow with her head pressed against the animal's side for balance. The milk pail is set on the ground or held against the milker's thigh by a hook attached to her belt and fitted into a notch in the pail. Robert B. Ekvall, who had extensive experience among the Tibetans of northeastern Tibet, has described the hook as a prized item of female

attire that, when wealth permits, is covered in silver, coral, and turquoise. According to Ekvall, the milking hook has come to symbolize, in nomadic society at least, the female's role in foodgetting (Ekvall 1968: 50).

No distinction is made between milk taken in the morning and that collected in the evening; both are considered equal in all respects. For children, especially, milk is considered a very healthy and nutritious food and may be made more appetizing by the addition of sugar, butter, or some other ingredient. The Mewu Fantzu, a pastoral group of Kansu, are said to consider yak milk better for children than even mother's milk, which they say makes children stupid (Stübel 1958: 8). Although this is probably a radical position, it nonetheless serves as an indicator of the esteem in which yak milk is held.

While some Tibetans allow small children to drink fresh milk, it does not seem to be the practice among adults or children over 4 or 5 years of age. For them, milk must be boiled before it is consumed. When asked to explain why they boil milk before drinking it, many Tibetans insist it is necessary to avoid the diarrhea, flatulence, vomiting, and stomach cramps that would surely follow the drinking of fresh milk. Some, however, maintain that milk is boiled simply to improve its taste and that no medical considerations are involved. Whatever the reason, milk is not boiled very long but rather is heated just until it boils, after which it is quickly removed from the fire.

Milk Products: Butter and Cheese

Most milk is converted into milk products as quickly as possible. There are two reasons for this practice. First, milk processing results in products much less perishable than liquid milk. Second, such products can be transported more easily than liquid milk, even over short distances. Because a great deal of milk is produced and collected in pastures distant from established villages and milk yields vary from season to season, transportability and perishability are important factors that must be dealt with by pastoralists.

Butter is, perhaps, the most important product derived from milk both in amount produced and quantity consumed. Butter derived from the milk of yak has the highest prestige value and is preferred over all others for its taste, richness, and color, which varies from season to season from a golden hue in summer to pale yellow in late winter. No distinction seems to be drawn among butters of various hybrids.

The major instrument of butter making is the churn, and a variety of types are employed, with some made of wood, some of leather, and others of pottery.⁸

Butter making, like milking, is principally a female occupation. However, no universal rule exists on this point. As with milking, the process of butter making varies little. After milk is collected from animals, it is warmed over a fire and stored for about a day perhaps in a churn, or in vessels assigned for that use. In either case, the warmed milk is allowed to curdle. Some-

times, a small amount of starter is added to hasten curdling. Once curd is formed, the mass is transferred to a churn and agitated. As soon as butter begins to form, a quantity of warm water is added a little at a time, to encourage the process. As larger clots of butter float to the surface of the buttermilk, they are picked out and squeezed by hand to remove any remaining liquid.

Those who maintain large herds of milch cows produce more butter than is necessary for the normal needs of their families; thus, they store excess butter in a variety of ways for future use, sale, and trade. A common method of storing butter in the Nepal Himalayas nowadays is simply to pack it solidly in 5-gallon kerosene tins. Before tins were commonly available, however, butter was packed in sheep stomachs or wrapped in wet rawhide bags, which, sewn shut and dried, are said to have kept butter from a few months to a year or more. Butter is also stored in wooden boxes and barrels and, according to some Tibetans, pottery containers as well.

Butter, a valuable low-bulk manufacture, is traded widely and is sold by weight. Demand for butter is the result of various factors, not least of which is its prominent role in the diet. Seldom is butter eaten in its raw state. More often it is melted and used in frying or mixed with other foodstuffs. The largest percentage of butter is consumed in the form of *tsocha*, the ubiquitous butter-tea of Tibet and the Himalayan region. *Tsocha* is souplike in consistency and nourishment. And, although the ingredients do not vary at all from place to place, the taste of *tsocha* varies from family to family, perhaps because of the quality or quantity of ingredients used.

Only three ingredients are necessary to make *tsocha*: tea, butter, and salt. Tea, even among the poorest, must be Chinese, imported at great expense in brick or cone form. A quantity of tea is broken off and boiled in water that is then strained and poured into a tea churn. After hot tea is poured into the churn, butter and salt are added. The best *tsocha* is said to be made with yak butter; when this is unavailable, butter of hybrids is used, followed, in order of preference, by the butters of common cows, sheep, and goats. Salt is almost always of Tibetan origin; Indian salt is seldom used because it is thought to cause illness. After being thoroughly churned, the *tsocha* is poured into a kettle for reheating; when hot, it is served in handleless cups or in glasses. *Tsocha* is a dietary staple drunk many times a day by persons of all ages.

Two other factors help explain the high value of butter and the constant demand for this commodity in the marketplaces of Tibet and Nepal. First is the role of butter in discharging social obligations, and second is the multifarious ceremonial role of butter.

With respect to social obligations, in many instances payment in the form of butter is a desirable method of settling accounts. Turning briefly to religious and ceremonial requirements, we find an almost insatiable demand for butter. In lamaserie, for exam-

ple, large quantities are required for making the *tsocha* served to lamas, and enormous amounts are consumed in the many votive lamps kept burning on altars. One scholar has suggested that the larger lamaserie of Tibet required tons of butter each year for just these two purposes (Tucci 1967: 94).

Unlike butter, however, there is no great market for or monetary value attached to the soft and hard cheeses made from yak milk. Nonetheless, cheese, made from buttermilk, is appreciated as a nourishing and tasty addition to meals; as an easily carried, sustaining food for long journeys; and as a treat on festive occasions.

Women, who are solely responsible for cheese making, do not hesitate to mix the buttermilk of yak, hybrids, and common cows in order to produce cheese. The principal concern seems to be the availability of enough buttermilk to warrant the long, tedious process of cheese making and not the composition of the buttermilk.

In any case, after the butter is removed from a churn, the remaining buttermilk is poured into kettles which are set over a fire. The buttermilk is brought to a rolling boil, after which it is removed from the fire to cool. As it cools, the buttermilk separates, forming a thick, spongy, white mass that floats to the surface of a clear liquid.

Solids are skimmed off with a shallow strainer, something like a dipper, and are placed in a cloth bag to allow whey to drain off. What remains in the cloth constitutes soft cheese, which is eaten immediately. On special occasions, sugar is added to make a sweet, tasty mixture. Soft cheese is also fried before eating, mixed with barley flour to form a doughlike food, or boiled with butter, salt, and other spices and served as a soup. More commonly, however, soft cheese is dried to form the second, hard type of cheese.

Hard cheese is produced by drying soft cheese in the sun or by the fire. Sun-drying is much preferred over fire-drying, for the latter darkens the cheese and makes it less tasty. If weather permits, soft cheese is scattered on cloths spread on the ground close to the house. If this is not possible, soft cheese is arranged before an indoor fire, on mats woven of vegetable material. In either case, soft cheese shrinks to about a fourth of its original size, forming a very hard cheese that is said to keep indefinitely.

Hard cheese most commonly is eaten by slowly dissolving pieces in the mouth. But this kind of cheese is also added to stews and soups or dropped into cups of *tsocha*. As an easily carried food, cubes of hard cheese are often taken on journeys to supplement meals purchased on the trail, or they may be the sole sustenance of the traveler.

Blood

Bleeding of live cattle is a widespread and common practice among the peoples of Tibet and Nepal. Bleeding is done for a variety of purposes, principal among them to obtain blood for human consumption.

Unlike milk, butter, and cheese (which are available and eaten throughout the year), blood is a seasonal food. Bleeding of yak for food takes place in late spring and early summer; only if there should be a famine are cattle bled for food in other seasons.

Not all yak are bled for food, however. Yak oxen are the animals most often bled, but occasionally yak cows, if they are not lactating, are bled as well. Hybrid cows seem never to be bled for food, and only persons who have no yak, or insufficient numbers of them, bleed male hybrids. There seems to be no evidence of bleeding common cattle for the sole purpose of procuring blood for food, though they are bled for other reasons. Moreover, no consistent and uniform explanation exists for not bleeding yak dams, female hybrids, and common cattle, but the strong impression is that although bleeding males for food is not quite proper, bleeding females for food is quite offensive.

Perhaps because bleeding cattle for food offends Buddhist sensibilities, Tibetans and Himalayans derive most blood food as a by-product of bleeding for other purposes. One is to prevent disease in cattle and ensure the cattle's survival by encouraging them to gain weight quickly.⁹ Such operations, like bloodletting for food, generally take place in late spring and early summer when cattle, after surviving the winter, begin feeding on new grass.

Any animal considered lean or prone to disease is a candidate for bleeding, though more time and consideration is given to the selection of milch cows and dams lest bloodletting cause lactation to cease. Even calves are bled to improve their health and chances for survival. Cattle that are sick or so weak that they are in immediate danger of dying are also bled. But blood drawn from such animals is not considered fit for human consumption.

As previously noted, lactating yak dams and hybrid cows are seldom, if ever, bled to supply their owners with food. And a great deal more care is shown in the selection of females for prophylactic bloodletting operations than is extended to males. Perhaps this more sympathetic treatment of cows is, as some Tibetans suggest, an attempt to reconcile human food needs with the Buddhist view that bloodletting is an improper activity because it involves use of a knife and the spilling of blood.

Or perhaps the explanation of more secular Tibetans is correct. According to them, milch cows and dams are rarely bled because there is always a possibility that the operation will prove fatal or result in cessation of lactation. This thinking suggests that decisions are based purely on selfish and economic grounds and not on higher concerns. In any case, no such consideration is shown to dry cows and to those that have difficulty being impregnated. These animals are bled in spring to improve their chances of pregnancy.¹⁰

Whatever the purposes of bleeding, the method

varies little from place to place and from group to group. Essentially, bleeding begins with the application of a tourniquet, simply a length of rope, around the animal's neck. Often, but not always, a similar tourniquet is tied around the animal's belly or buttocks to force (it is maintained) blood to the neck. Once in place, the neck tourniquet is tightened. If accompanied by another tourniquet, both are tightened. This action causes the blood vessels of the neck to stand out. One of these vessels is chosen, and using a small awl-like instrument or a special handleless chisel-edged tool called a *tsakpu*, the bloodletter makes an incision (not a puncture) of about 1/4-inch to 1/2-inch in length. The blood that spurts from the wound is caught in a container if it is to be prepared as food, or it is allowed to fall directly to the ground if considered inedible.

Varying quantities of blood are taken from cattle, depending primarily upon the sex, size, and condition of the animal. If the animal is male and in good condition, a half gallon of blood may be taken; if it is female and in poor condition, only a pint or so may be removed. If the operation is performed correctly, bleeding stops as soon as the tourniquet is loosened.

Bleeding of cattle is strictly a male task; women never perform the operation under any circumstances. Among nomads, each adult male is an accomplished bloodletter, having learned from his elders during childhood. Among settled folk, some perform the operation on their own cattle, but others, not as skilled, often prefer to have more experienced men bleed their cattle.

If bleeding of cattle is a task assigned to men, preparation of blood for eating is strictly a female occupation. Blood food must undergo some kind of processing because nobody would consume fresh blood. There are a number of methods of preparing blood drawn from live cattle; the particular one employed by a family depends entirely upon personal preference. One of the most common ways of preparing blood food involves simply allowing the blood to coagulate (a process that sometimes is hastened by the addition of salt), after which it is either boiled in water or fried in butter.

Another method of preparation requires that the blood be mixed with water; after a period of time, a thick mass forms that is eaten plain or mixed with salt and/or barley flour. Or barley flour may be mixed with fresh blood to form a dough that is then cooked. Some prefer to put fresh blood into metal containers that are set over a fire to warm. When the blood coagulates into a jellylike substance, it is seasoned with salt and cut into cubes to be eaten plain, fried, or boiled.

Finally, there is a blood food that seems to be peculiar to the Sherpa of Solukhumbu in Nepal. Blood drawn from cattle is allowed to dry out thoroughly and is rubbed between the hands to form a powder. This powder, or "blood flour," is added to various dishes as a seasoning.

However it is prepared, blood is not considered entirely proper as a food. Some Tibetans and Himalayans, more committed to the Buddhist concept of nonviolence, believe that the consumption of blood food is sinful even though the process of blood-letting may benefit cattle. Nevertheless, virtually everyone eats blood food with relish, finding it tasty and nourishing. This religious-dietary dilemma is, perhaps, understandable when it is remembered that blood food is a seasonal commodity, available in large quantities during late spring and early summer. This timing coincides with the period of meat shortage for most Tibetan and Himalayan folk. Meat derived from stocks put up the previous autumn is near depletion at this time, and blood is a welcome addition to the diet, too tempting to forgo.

Richard P. Palmieri

The author is grateful to the National Geographic Society for two research grants that supported fieldwork in the Nepal Himalayas in 1972 and 1973, and to Mary Washington College for financial support that enabled preparation of this manuscript.

Notes

1. Works that focus on yak or at least deal with this animal in a more than cursory fashion include the recent studies of the Sherpa of Khumbu in Nepal by Barbara A. Brower (1991) and Stanley F. Stevens (1993), the collection of essays in the volume edited by Joseph Bonnemaire and Corneille Jest (1976), the several articles by Richard P. Palmieri (1972, 1974, 1980, 1982, and 1987), and three books on Tibetan pastoral nomadism: one by Matthias Hermanns (1948), another by Robert B. Ekvall (1968), and the third by Melvyn C. Goldstein and Cynthia M. Beall (1990).
2. It is, perhaps, useful to clarify the term "yak." It is Tibetan in origin (Tib. *gyag*) and refers specifically to the domesticated male. Females are correctly known as *di* or *dri*; in some regions of the Himalaya, however, females are also called *nak*, although this term more precisely refers to cattle in general.
3. The question of yak hybridization is the focus of several papers, many of them referred to in my study (Palmieri 1987), which deals with the complex nomenclature, animal husbandry practices, and ecology of hybrids among the Sherpa of Nepal.
4. A detailed discussion of the concept of pollution among the Sherpa, for example, can be found in Ortner (1973).
5. The literature on this topic includes a wide range of conflicting figures that are difficult to reconcile. This variability may simply represent data collected in different environments, the contrasting characteristics of dissimilar herds, or the various seasons during which the figures were developed. In any case, the following sources may be consulted: Das (1904: 96-7); King (1926: 75); Kislovsky (1938: 32); Phillips, Johnson, and Moyer (1945: 70); Phillips, Tolstoy, and Johnson (1946: 170); White, Phillips, and Elting (1946: 357); Hermanns (1948: 128); Stübel (1958: 19); Fürer-Haimendorf (1964: 12); and Field and Pandey (1969: 50-2).
6. A discussion of the biological and behavioral bases for the milk let-down reflex can be found in Amoroso and Jewell (1963).
7. In this last, it should be noted that African and European parallels do exist. See, for example, Lagercrantz (1950: 44-50) and Bühler-Oppenheim (1948: 4275-6). With respect to insufflation of yak, see Edgar (1924: 39); Combe (1926: 105-6); Bühler-Oppenheim (1948: 4276); Duncan (1964: 240); and Ekvall (1968: 50).
8. The principal difference between various wooden churns is the apparatus employed to agitate the milk: Some are equipped with dashers, others with paddle-like arrangements. Both the pottery and leather churns are rocked or shaken to produce butter.
9. Some Africans bleed their cattle for similar reasons. See, for example, Kroll (1928: 244) and Lagercrantz (1950: 51-4).
10. Interestingly, farmers in the Bavarian and Austrian Alps also bleed their barren cows just before rutting season in an attempt to increase their chances of becoming gravid. For this information, I am most grateful to Professor Ludwig Erhard of the Tierzucht Forschung e. F. München Institut für Blutgruppenforschung (Munich).

Bibliography

- Amoroso, E. C., and P. A. Jewell. 1963. The exploitation of the milk ejection reflex by primitive peoples. In *Man and cattle: Proceedings of a symposium on domestication*, ed. A. E. Mourant and F. E. Zeuner, 126-37. London.
- Bohlken, Herwart. 1958. Zur Nomenklatur der Haustiere. *Zoologischer Anzeiger* 160: 167-8.
- Bohlken, Herwart. 1958-60. Vergleichende Untersuchungen an Wildrindern (Tribus Bovini Simpson 1945). *Zoologische Jahrbücher* 68: 113-202.
- Bonnemaire, Joseph, and Corneille Jest. 1976. *Le yak: Son rôle dans la vie matérielle et culturelle des éleveurs d'Asie centrale*. Paris.
- Brower, Barbara A. 1991. *Sherpa of Khumbu: People, livestock, and landscape*. Delhi.
- Bühler-Oppenheim, Kristin. 1948. Hilfsmittel zur Erleichterung des Melkens bei Naturvölkern. *Ciba Zeitschrift* 10: 4275-6.
- Combe, G. A. 1926. *A Tibetan on Tibet: Being the travels and observations of Mr. Paul Sherap (Dorje Zödba) of Tachienlu*. London.
- Das, Sarat Chandra. 1904. *Journey to Lhasa and central Tibet*. New York.
- Duncan, Marion H. 1964. *Customs and superstitions of Tibetans*. London.
- Edgar, J. Huston. 1924. A strange milking custom. *Man* 24: 39.
- Ekvall, Robert B. 1968. *Fields on the hoof: Nexus of Tibetan pastoral nomadism*. New York.
- Field, D. I., and K. R. Pandey. 1969. *Pasture, fodder, and livestock development: Trisuli watershed - Nepal*. Kathmandu, Nepal.
- Fürer-Haimendorf, Christoph von. 1964. *The Sherpas of Nepal: Buddhist biglanders*. Berkeley, Calif.
1966. Caste concepts and status distinctions in Buddhist communities of western Nepal. In *Caste studies in Hindu-Buddhist contact zones*, ed. Christoph von Fürer-Haimendorf, 140-60. New York.
- Goldstein, Melvyn C., and Cynthia M. Beall. 1990. *Nomads of western Tibet: The survival of a way of life*. Berkeley, Calif.

- Hermanns, Matthias. 1948. The A Mdo Pa Greater Tibetans: The socio-economic bases of the pastoral cultures of Inner Asia. Ph.D. thesis, Universität Freiberg in der Schweiz, translated for the Human Relations Area Files by Frieda Schutze. New Haven, Conn.
- King, Louise (Rin-chen Lha-mo). 1926. *We Tibetans*. London.
- Kislovsky, D. 1938. The domesticated animals of Mongolia. *Journal of Heredity* 29: 27-32.
- Kroll, Hubert. 1928. Die Haustiere der Bantu. *Zeitschrift für Ethnologie* 60: 177-290.
- Lagercrantz, Sture. 1950. *Contribution to the ethnography of Africa*. Lund, Sweden.
- Mittaine, Jean. 1962. Milk other than cows' milk. In *Milk hygiene*, ed. M. Abdussalam, et al., 681-94. World Health Organization Monograph Series No. 48. Geneva.
- Ortner, Sherry B. 1973. Sherpa purity. *American Anthropologist* 75: 49-63.
- Palmieri, Richard P. 1972. The domestication, exploitation, and social functions of the yak in Tibet and adjoining areas. *Proceedings of the Association of American Geographers* 4: 80-3.
1974. Culture and ecology of Tibetans in the Greater Nepal Himalaya. In *The Conference on Nepal at Claremont*, ed. Merrill R. Goodall, 4-5. Claremont, Calif.
1980. Preliminary results of field investigations in the Nepal Himalaya. In *National Geographic Society research reports*, ed. Paul H. Oehser, John S. Lea, and Nancy Link Powars, 529-34. Washington, D.C.
1982. Patterns of Indian pastoralism. In *India: Cultural patterns and processes*, ed. Allen G. Noble and Ashok K. Dutt, 325-36. Boulder, Colo.
1987. Cattle hybrids among the Sherpa of Nepal. *Journal of Cultural Geography* 7: 89-100.
- Phillips, Ralph W., Ray G. Johnson, and Raymond T. Moyer. 1945. *The livestock of China*. Washington, D.C.
- Phillips, Ralph W., Ilia A. Tolstoy, and Raymond G. Johnson. 1946. Yaks and yak cattle hybrids in Asia. *Journal of Heredity* 37: 162-70, 206-15.
- Schley, Peter. 1967. *Der Yak und seine Kreuzung mit dem Rind in der Sowjetunion*. Wiesbaden.
- Sherring, Charles A. 1906. *Western Tibet and the British borderlands*. London.
- Stevens, Stanley F. 1993. *Claiming the high ground: Sherpas, subsistence, and environmental change in the highest Himalaya*. Berkeley, Calif.
- Stübel, Hans. 1958. *The Mewu Fantzu: A Tibetan tribe of Kansu*, trans. Frieda Schutze. New Haven, Conn.
- Tucci, Giuseppe. 1967. *Tibet: Land of snows*, trans. J. E. S. Driver. New York.
- White, W. T., Ralph W. Phillips, and E. C. Elting. 1946. Yaks and yak-cattle hybrids in Alaska. *Journal of Heredity* 37: 354-8.



PART III

Dietary Liquids

Drinks, many of which were also globalized in the events following 1492, seem to have produced even more ill effects than foods. Unfortunately, more often than not, they did indigenous peoples unaccustomed to them no good. Along with wine and distilled beverages, beer – which had been brewed since the early stages of the Neolithic – entered all newly discovered corners of the earth as quickly as the Europeans reached them. Superficially more healthful was the spread of tea from China, coffee from Ethiopia, and cocoa from the Americas. But consumers of these beverages also demanded the addition of sugar and the processing of this Asian plant soon became an American enterprise fueled by unwilling African labor, which was also applied to coffee production.

Water that fostered disease was a problem from the very beginning of the Neolithic, and one that grew steadily worse as more and more people crowded together to pollute local sources. Alcoholic beverages were humankind's first line of defense against pathogen-packed water, and later, the procedure of boiling water to make tea, coffee, or cocoa had the same purifying effect. But it was only in the twentieth century that public water supplies were rendered safe to drink in the world's developed countries, and unsafe water remains a problem throughout the developing world.

This means, among other things, that formula feeding of the young continues to cause infant deaths in that world when local water is used, making breast milk by far the best alternative (save of course in the presence of AIDS). Bovine milk, by contrast, is not an especially good food for infants and is positively dangerous for some. In the past, bovine milk imparted bovine tuberculosis (scrofula) to the lactose *tolerant*, and, of course, milk drinking did not even catch on among the bulk of the planet's people (who are lactose *intolerant*), although it is the case that many reduced-lactose products of milk, such as yoghurt and cheeses, have become welcome dietary additions.

Caffeine-carrying kola nuts from Africa helped to launch the soft-drink industry, which became a giant all across the globe during the course of the twentieth century and provided even more stimulation for the sugar industry, not to mention the dental profession.

Khat and kava are among the few beverages of the world that have not been globalized. Khat (more often chewed these days than made into a beverage) remains confined to northeastern Africa, whereas kava is a drink of the Pacific. In the past confined to ceremonial affairs, kava, with its narcoticlike properties, has become a local alternative to alcohol.

III.1 🍷 Beer and Ale

Beer and ale are mildly alcoholic beverages made from the action of yeast fermenting a usually grain-based mixture. Throughout their history, they have constituted both a refreshing social drink and an important energy-rich food. The basic ingredients of most beers and ales have included grain, water, yeast, and (more recently) hops, and despite many regional variations, the process of fermenting the grain has changed little over time. To be completely accurate, it must be noted that ale is defined as unhopped beer; in this chapter, however, the terms “beer” and “ale” are employed interchangeably for the period before hops were used.

The Chemical Basis of Fermentation

Before fermentation can take place, yeast, a single-cell fungus occurring naturally in several varieties, must be allowed to act on the sugar present in grain. This releases two crucial by-products, alcohol and carbon dioxide. A grain often used for this purpose is barley – even though, in its natural state, it contains only a trace amount of free sugar – because of its high content of starch, a complex polymer of sugar. Barley also contains substances known collectively as diastases, which convert the barley starches into sugar to be used as food for the growing plant. When barley is crushed and dried carefully, the essential starches and diastases are released and preserved, rendering a substance called “malt.”

Until sometime around the ninth century, “beer” was actually “ale,” made by a process known as mashing, whereby the barley malt was mixed with hot – but not boiling – water. The effect of the hot water was to induce the diastases to act immediately in breaking down the complex starches into sugar. This process is referred to as conversion and results in “wort,” one of its most essential products. The mashing procedure not only produced the brown, sugary wort but also permitted inert elements of the barley, such as the husks, to be drawn off. In the production of pure ale (such as the first human brewers would have made), all that remained was for yeast to act upon the wort so that the sugars could be converted into alcohol and carbon dioxide.

Beginning in approximately the ninth century in central Europe (authorities vary widely and wildly regarding the date but not the place), the procedure began to be modified, and beer came into being with the addition of blossoms from the hop plant. Numerous modern beers are labeled as ale, but as mentioned, technically “ale” means unhopped beer. In order to convert ale into beer, dried hop blossoms are added to the boiling wort mix after the mashing but before the yeast is allowed to act. This releases two resins, lupulon and humulon, that act as excellent natural preservatives, preventing the growth of certain

types of bacteria which, although harmless to humans, are detrimental to beer. Before the use of hops, pure ale had a very limited “shelf life” and often spoiled, much as milk does. The diastases in barley also acted against the bacteria in question, but not nearly so effectively as hops. In fact, it can be argued that it was the harnessing of the preservative power of hops that permitted the production, storage, and distribution of beer in large quantities. Moreover, in addition to its antibacterial properties, the hop plant adds flavorful oils that mask the otherwise sweet taste of pure ale (Kloss 1959: 31–2).

Varieties

Today, the use of hops is standard, and very little pure ale has been mass-produced in the twentieth century. Thus, the words *ale* and *beer* have become largely (if wrongly) synonymous. Modern technology and advanced techniques have modified and refined the brewing process considerably. The most commonly mass-produced type of beer is known as a “lager” or a “Pilsner” and is lighter in color and generally milder in taste than other beers. The vast majority of North and South American beers, most European beers, the beers of Australia, and those of nearly all major Asian nations are crafted in the Pilsner style.

Darker beers that are dryer and richer in taste are referred to as porters and stouts, with the latter merely a stronger, drier porter. Several popular German beers make use of large quantities of wheat rather than barley to make *Weizenbier* (wheat beer). Belgium is famous for fruity ales, typically known as “lambic” beers, the production of which involves a complex process of spontaneous fermentation. In addition to the use of fruits to add flavor, as in the case of the lambic beers, one method of increasing taste is “dry-hopping,” a process whereby additional hops are added at the end of the process to replace the residue lost when the wort and hop blossoms are first boiled together. Like hops, yeast and sugar are occasionally added to the bottle when it is sealed to further enhance the beer’s strength. This is not so much the production of yet another style of beer as it is a method of setting up a secondary fermentation process within the bottle to make stout, porter, and “bitter.”

Earliest Origins

No one has yet managed to date the origins of beer with any precision, and it is probably an impossible task. Indeed, there are scholars who have theorized that a taste for ale prompted the beginning of agriculture, in which case humans have been brewing for some 10,000 years (Katz and Voigt 1986). Most archaeological evidence, however, suggests that fermentation was being used in one manner or another by around 4000 to 3500 B.C. Some of this evidence – from an ancient Mesopotamian trading outpost called

Godin Tepe in present-day Iran – indicates that barley was being fermented at that location around 3500 B.C. Additional evidence recovered at Hacinebi Tepe (a similar site in southern Turkey) also suggests that ancient Mesopotamians were fermenting barley at a very early date (Smith 1995).

At present, however, the evidence from both sites is sufficiently sparse to preclude any definitive assertion that the ancient Mesopotamians were the first people to make beer. On the other hand, one can speculate: There is no question that fermentation takes place accidentally (as it must have done countless times before humans learned something about controlling the process), and most investigators believe that barley was first cultivated in the Fertile Crescent region of lower Mesopotamia between the Tigris and Euphrates rivers. Grain is heavy to transport relative to the beer made from it, so it is not surprising that there may be evidence of ale in these outposts and not unreasonable to suspect that accidental fermentation did occur at some point in the ancient Mesopotamian region, leading to beer making (Corran 1975: 15–16).

In any event, we know that not much later the Sumerians were, in fact, making beer. The clay tablets (unearthed from the ancient city of Uruk in Lower Mesopotamia – now Iraq – and dating from the second half of the fourth millennium) that tell the story of Gilgamesh, the fifth king of the second Sumerian dynasty, make it clear that ale was in widespread use (Toussaint-Samat 1992), and Reay Tannahill (1988: 48) has written that “a staggering amount of the Sumerian grain yield went into ale; something like 40 percent of the total.”

At approximately the same time, people of the ancient Nubian culture to the south of Egypt were also fermenting a crude, ale-like beverage known as *bousa*, which is still brewed by African farmers and peasants to this day (Smith 1995). Indeed, although many scholars maintain that the Mesopotamian culture was the first to brew beer, others argue that it was the ancient people of East Africa who first produced and consumed a fermented product (Dirar 1993: 20).

In much the same fashion as with grain, the fermentation of fruit and fruit juices probably also occurred by accident at around this same time, leading to the earliest forms of wine. What is more difficult to ascertain, however, is how much knowledge ancient people had of the process. It is also difficult to know with any reasonable certainty how extensive their use of fermented barley was and exactly how much their ale might have resembled what the world now recognizes as beer.

The Importance of Ale in Early Societies

From the beginning of its production, ale (even in its crudest forms) would have been an important addition to an otherwise frequently limited diet. Resem-

bling the chemical makeup of bread in several ways, ale was a convenient package of starches, sugars, and other grain by-products that provided nutritional supplementation. Similarly, for people with few means of storing foods for any length of time and who depended on the vagaries of nature for subsistence, ale could be an excellent (and doubtless at times vital) source of calories.

Moreover, ale (and later beer) afforded an escape from the feces-fouled drinking water that plagued peoples for millennia. Although humans, until very recently, had no knowledge of pathogenic infection, water (and milk) was understood to provoke dangerous illnesses, even death, whereas fermented beverages were considered safe. And because of sterilization by boiling and by the action of yeast, this was generally the case.

Ale was also important because the earliest cultures, particularly those of the Sumerians and Egyptians, attached religious significance to its consumption. And throughout the ages, savants frequently maintained that ale had curative properties. But probably the most important reason for drinking ale and other alcoholic beverages was to achieve a desired level of intoxication. Because invectives against drunkenness can be found in both the Bible and the Koran, we know that people, beset by life's hardships or just seeking relaxation, were reaching that goal a long time ago. Indeed, the ancient Egyptians are credited with celebrating ale consumption by composing some of the world's earliest-known drinking songs.

Brewing in Antiquity

Although the fermenting of barley probably developed independently in several cultures, knowledge of brewing technology doubtless was spread throughout the Middle East by various nomadic peoples. One aspect of brewing technology common to the Egyptians, Sumerians, and Babylonians alike was the use of baked loaves of malted barley and grain that resembled baked bread. There were several variations of this technique, but basically the loaves of barley and wheat, once baked, were covered with water to form a mash, which was then placed in an earthen vessel for a time. In some cases, fermentation probably occurred spontaneously. In others, it was doubtless generated by the presence of yeast cells in the cracks and linings of the earthen vessels that were used over and over again. But in addition, skillful brewers had most likely learned to keep the remains of a previous mix to use as a starter (Smith 1995: 12–13).

The Greeks probably gained most of their understanding of brewing from the Egyptians, although the Babylonians may also have passed along what they knew. The Roman Empire, at its height, had the advantage of being able to borrow brewing techniques eclectically from several cultures. Roman historians, for example, did not credit Rome with spreading

information on ale making to the Germanic tribes of Europe. Rather, Tacitus recorded that these peoples were already fermenting a beverage from barley or wheat when they came into contact with Rome. Pliny also wrote of the barbarians and their beer, and it seems likely that the tribes of central and northern Europe gained brewing knowledge not from the Romans but from the Babylonians and other Asian civilizations. Or it could have been a situation such as that of the Celts of the British Isles, who are said to have developed a crude process of fermentation independently, but refined their ale-making skills with technology from other cultures (Corran 1975: 23–4).

Brewing in the Islamic World

By the time of the collapse of the Roman empire in the fifth century, the production of alcoholic beverages had been expanded, and beer was only one of many alcoholic beverages produced in the Arabian peninsula – a list which included a honey-based mead and fermented camel's milk. Their consumption – and especially that of ale – was widespread before the advent of Islam, despite a number of localized religions that had instituted prohibitions against it. Along the caravan and trading routes, houses, taverns, and inns were prosperous businesses that supplied beer and mead to travelers and, in some locations, to townspeople as well (Ghalioungui 1979: 13–15).

The spread of the Islamic religion did not, at first, bring restriction of alcoholic beverages; indeed, the Koran, like the Bible, celebrated the drinking of wine. Rather quickly, however, Islamic teaching began to forbid drinking alcohol, although the degree to which the rule was observed varied from place to place (Ghalioungui 1979: 15).

Egypt was one area in which alcohol continued to be used, although in the years following the entrance of Islam into Egypt, various rulers periodically enforced the Muslim prohibition. But the consumption of fermented beverages was never entirely eradicated (Ghalioungui 1979: 15), and, among the peasant population, *bousa* continued to be produced and consumed as it had been for centuries.

Despite such exceptions, however, Islam had an enormous impact on beer brewing in the Middle East and elsewhere in the Muslim world, with the result that it never was the mass-produced, socially accepted beverage that it became in Europe during the Middle Ages. Europeans, especially monks in the monasteries of the Catholic Church, not only kept the knowledge of brewing alive but also began its refinement into a modern science.

Brewing in Europe

Over the course of the Middle Ages, beer brewing flourished in northern Europe (where foods laden with carbohydrates and fats required much liquid to

wash them down) and evolved into a distinct industry (Tannahill 1988). As such, by the end of the Middle Ages, beer had become subject to taxation and also to government regulation (especially in Britain and the German states) aimed at standardizing the brewing process.

In the early Middle Ages, however, monasteries and churches were the principal ale makers in Europe. Because the church was at the center of the lives of the people, monasteries and churches commonly provided the settings for festivals, weddings, and other social gatherings that were lubricated with ale. Indeed, such was the control of the church over access to ale that it became a device for ensuring the participation of parishioners in church rituals. Later, as guilds developed, the church influenced – even controlled – many of their activities with the promise of ale.

In addition, monasteries were much more than just monastic retreats. The growing of food was one of the monks' primary occupations, and as a rule, the land owned by their orders was sufficient enough for the rotation of crops in such a way as to ensure a constant supply of cereals. Much of the cereal produced – including spelt, wheat, oat, and rye, as well as barley – went into ale, the quality as well as quantity of which the monks improved upon, just as they did with their wines and cheeses (Toussaint-Samat 1992). Many monasteries also served as inns that provided room and board for travelers, and some became famous for their ales, their praise carried by church pilgrims, merchants, and others on the move. There is no question that monastery-produced ales, made on a near-industrial scale, brought in a very good income for the various orders.

Later, however, as the Middle Ages progressed, ale production in the towns and countryside began to rival that of the church. And as the craft passed into private hands, it mirrored other early trades with its guilds and specialization. Because of its limited shelf life (prior to the use of hops), ale was usually brewed and distributed on the same site, and consequently, the first brewers outside of the church were generally boardinghouse owners and tavern keepers who provided ale to travelers and guests. Local inns and taverns came to be regarded by townspeople and villagers alike as social gathering places (Corran 1975: 36–7).

Because water is vital to the brewing process, the breweries of taverns and inns had to be located near an abundant water supply. But the type of water was important. If it was hard water with lime, the fermentation process might not work well; if it had iron in it, the beer would always be cloudy.

Women frequently oversaw the breweries while their husbands ran the inns. In fact, women were much involved in the ale business, sometimes owning boardinghouses as well as breweries and holding special licenses to distribute their product.

By the end of the Middle Ages, breweries and drinking establishments of one sort or another had multiplied to the point where they overshadowed the monasteries, both in England and on the Continent. As the church ceased to dominate the brewing industry, states began to take an interest in both taxing and (because of increasing adulteration) regulating it. An example of the former is the 1551 licensing of English and Welsh alehouses for the first time (Trager 1995), although the classic example of regulation had come earlier, in 1516, when William VI, Duke of Bavaria, instituted a *Reinheitsgebot* – an “Edict of Purity” – which decreed that the only ingredients permitted in beer were water, barley, malt, yeast, and hops. The edict is still in effect, now for all of Germany, but it is said that only Bavaria holds to it for exported beers.

Hops, which converted ale into modern beer, were coming into widespread use at about this time. Hop blossoms are derived from the hop plant (a relative of *Cannabis*), and as their use became common, a hop garden was an essential component of a brewery. As noted, the aromatic hop greatly enhanced the taste of ale, as did the addition of other herbs and flavorings. More importantly, however, hops greatly extended the life of ale, which in turn removed the necessity for locating breweries and taverns close to one another. The use of hops was eventually so universal that the brewing of pure ale became nearly extinct, until the modern, twentieth-century “Campaign for Real Ale” movement in Britain sought to revive what was perceived as a dying art.

The revolution that hops worked in the brewing industry, however, was a long time in coming. Since Neolithic times, hops were believed good for one’s health and sometimes carried the burden of a reputation as an aphrodisiac. It has been suggested that the utilization of hops in beer can be traced back as far as the ancient Egyptians. But we hear nothing of hops in beer in the Roman world. Pliny tells us only that the Romans ate hop shoots much like asparagus. During the early Middle Ages, hops were grown for medicinal purposes in gardens throughout the central European region from the North Sea and the Baltic to western Austria and northern Italy, but people apparently began putting them in ale only around the eighth or ninth century (Toussaint-Samat 1992).

The hop was only one of many herbs added to ales, but brewers sooner or later noticed that this herb improved the appearance of ale, that it acted as a diuretic, and most importantly, that it was a preservative. Nonetheless, the church successfully fought the widespread adoption of hops for centuries – apparently in part because of the aphrodisiac reputation of the plant, but also because the church, with its virtual monopoly on ale, did not welcome change. Moreover, hops were long viewed as an adulterant added to mask the taste of spoiled beer. Yet, somewhat ironically, it was probably the monks, with their considerable knowledge of medic-

inal herbs, who had added hops to ale in the first place (Toussaint-Samat 1992).

By the beginning of the sixteenth century, hops had become an essential ingredient for beers made on the Continent, and late in the reign of Henry VIII (died 1547), they were introduced to England. At first, the idea of adding hops to ale was a distressing one for the English, who continued to view them as adulterants and passed laws to prohibit their use. In 1554, however, Flemish hop growers emigrated to England to begin their production in Kent for a wary British brewing industry. Afterward, the use of hops was generally accepted, although many clung to their unhopped ales. Not until around 1700 was ale in England regularly hopped and the terms “ale” and “beer” accepted there as more or less identical (McGee 1984; Trager 1995).

The preservative powers of the hop plant contributed to the development of larger breweries producing beer in ever greater quantities – a trend in both England and Europe throughout the sixteenth and into the seventeenth centuries, especially in Flanders, northern and eastern France, and Bavaria, where the climate best suited the growing of hops. Because hops endowed beer with a greatly extended shelf life, brewers could now locate at a distance from the towns and, consequently, close to less-polluted stretches of streams and rivers, whose waters contributed to better-quality beers. Such moves were also necessitated by the regulatory measures of crowded cities, which sought to minimize the fire hazards arising from kilns burning in brewery buildings constructed of wood. Converting brewhouses into stone or brick structures or, alternatively, moving them out of the cities were both expensive options, and as a result, the sixteenth and seventeenth centuries in Europe also saw a trend toward fewer breweries – but much larger ones that, in many cases, were the forerunners of modern breweries still in operation at the end of the twentieth century.

With the mass production of high-quality beers, brewers cast an eye on the export market, and as exporting beer became a widespread endeavor, states enacted laws to regulate trade. One example of this trend comes from sixteenth-century England, where because of concern that the volume exportation of beer in wood casks and barrels would accelerate the dwindling of the island’s supply of timber, brewers were compelled to bring as much wood into the country as they sent away.

Brewing in the New World

Traditional interpretations hold that beer and brewing technology came to the Americas from Europe via the Jamestown settlers and the Pilgrims. However, several indigenous American cultures (outside of North America) had developed fermented products long before the Europeans arrived. In the Andean society of the

Incan empire, the fermentation of beverages was well established – the term *chicha* referring collectively to the numerous indigenous fermented beverages of South America.

The *chicha* of the Incas was elaborated primarily from maize, although there were variants, including beverages made from manioc roots and peanuts, to name just two. It is interesting to note that in the absence of hops, diastases (by-products important for flavoring and increasing alcohol content) were introduced to maize beer from human saliva, as moistened maize powder was rolled into balls and placed in the mouth (Morris 1979: 22).

Evidence from Spanish colonial sources and archaeological finds suggests that the production and consumption of maize beer was fairly widespread in the Andes area. Like the ales of Europe, *chicha* was not only a significant component of religious and economic life but served nutritional needs as well. Its importance was apparent in its large-scale, state-controlled production – revealed by archaeological excavations that have indicated the existence of state-run breweries and distribution centers (Morris 1979: 26–7).

The mass production of *chicha* in the Incan empire was abolished by the Spaniards, but the making of maize beer on a smaller scale remained widespread and continues today in the Andes and elsewhere. The relatively high price of modern beer makes *chicha* an attractive alternative in the rural areas of Central and South America.

Another indigenous American beverage that is still produced is a Brazilian beer known as *kaschiri*, which is fermented from manioc roots. Its manufacture is similar to the Incan maize beer in that the tubers are chewed and moistened by salivation. Maize is also used by Indians in Mexico to make a crude beer called *tesguino*. Far more pervasive, however, was *pulque*, the fermented juice of the *agave*, a plant which later was employed to make tequila. Like other local beverages among impoverished peoples, both *pulque* and *tesguino* still deliver important nutrients to their Mexican Indian consumers.

Since the Spanish conquest in Mexico and South America and the English settlement of North America, none of the indigenous American beverages have been produced on a large scale except for *pulque*, which remained a common drink of poor Mexicans until well into the 1940s. The mass-produced, twentieth-century beers of Central and South America are almost universally hopped Pilsners and employ techniques brought to the New World by Europeans.

Brewing in Early North America

Beer and ale were present from the beginning in the English settlements of North America. Records of both the London Company and the Jamestown colony indicate that beer reached the latter in 1607,

its very first year of existence. But in those early years, beer was too bulky (and thus too expensive) to transport efficiently; it also spoiled in the summer heat, and so the colonists soon began brewing their own (Rorabaugh 1979: 108–9). Although barley and hops were not at hand, other basic materials that would ferment, such as persimmons, pumpkins, maize, Jerusalem artichokes, and maple sugar, were abundant in eastern America, and by 1609, the governor of Virginia was advertising for brewers to come to the colony (Baron 1962: 4).

A bit later, in Massachusetts, the Puritans – like other Europeans of the age who justifiably viewed water consumption with intense suspicion – had followed suit and were brewing their own beer. The Puritans also pioneered some of America's first regulatory statutes for the production, distribution, and consumption of the beverage. By 1637, taverns and inns had to be licensed by the General Court, and it was forbidden for them to brew their own product. Rather, beer was to be obtained from a commercial brewer, also licensed by the court, who was enjoined to sell at court-specified prices. By 1629, similar regulations had also been adopted in Virginia (which now had two large brewhouses).

Not that there were all that many taverns in early America, and beer was more often than not brewed and consumed in the home. Those who had access to barley or could afford to import malt from England produced something a European might recognize; other beers continued to be made from local ingredients and from West Indies molasses. In addition, beer was imported on a fairly large scale from England, or from the Netherlands in the case of New Amsterdam, where the Dutch had established beer as a prominent drink back in the 1620s.

Beer production kept pace with the growing population throughout the remainder of the seventeenth century and into the eighteenth, with Philadelphia and Boston becoming major brewing towns. As was the case in Europe, the vast majority of colonial towns had taverns that not only provided places of lodging for travelers but also served as local social centers. These dispensed some beer, but rum and (to a lesser extent) corn whiskey and cider increasingly enjoyed more appeal, and even tea, made available in quantity by British mercantilism, cut into beer consumption. Nonetheless, that consumption grew anyway in the eighteenth century because the population was growing (Baron 1962: 56–8).

Innovations

The eighteenth century also saw numerous innovations in beer production, although as a rule, these were slow to reach America. One illustration has to do with what is typically called porter. During the period, working men in England would often order ale and a few other beers mixed together in a

tankard, and tavern keepers came to dread the extra time and effort required to satisfy such a request. Eventually, however, several of these brews were being mixed together in a single cask with extra hops. The resulting dark, strong, beverage was called porter, after the London laborers ("porters") who popularized it. It was during this period as well that many of the familiar English beers, such as Courage, Whitbread, and Guinness, were born.

Another significant technological innovation in beer brewing was glass bottling, which came into widespread use in the eighteenth century. Glass bottles made beer easier to transport and store and, after the advent of sealed bottles, extended its shelf life. But because the Industrial Revolution was first an English and European phenomenon (and because of British mercantile restrictions), glass bottling in America began in earnest only after the Revolution. Glass bottles enabled people to store and consume beer at home with greater ease and, perhaps coupled with the growth of alternative beverages, diminished the role of the tavern in the social life of towns.

By about 1750, coke and coal were providing maltsters with greater control over the roasting of malt, which made possible the brewing of pale and amber beers, and a classification system to differentiate these from the dark stout and porter brews became an important issue. Later in the century, thermometers and hydrometers added more control to the different stages of the brewing process, and in 1817, a "patent malt" was developed that made stout and porter brews lighter than they had previously been – beginning a trend toward less-alcoholic beer that continues to this day (McGee 1984).

Brewing in the Modern World

In the nineteenth century, beer brewing was revolutionized by a process that had originally been discovered back when hops were just beginning to find their way into ale on a regular basis. Until about 1400, "top" fermentation had been the procedure used. However, at about that time in Bavaria, the process of "bottom" fermentation was developed – "top" and "bottom" indicating where yeast collects in the vat. Bottom fermentation permits the manufacture of a lighter beer, but it was not until the 1840s that the technique spread from Bavaria first to Pilsen, Czechoslovakia, and Copenhagen, Denmark, and then to the wider world. "Pilsner lager" became the prototype of modern beers, with only England and Belgium persisting in the use of top fermentation.

Yet even before the spread of the new lager beer, the nineteenth century had begun to witness the rise of large-scale commercial breweries. These were encouraged by the growth of cities, which provided mass markets and rising wages for an ever-growing urban working class. By 1800, brewers in England, such as Whitbread and Barclay Perkins, were produc-

ing 100,000 to 200,000 barrels of beer per year. The largest brewer, Arthur Guinness, held a virtual monopoly in Ireland (Hawkins 1979: 14).

At about midcentury, German immigrants set about completely transforming the brewing industry in the United States. It was in 1844 that Frederick Lauer – a second-generation brewer in Reading, Pennsylvania – introduced the new lagering process, and the business of beer exploded. During the 1850s and 1860s, under the direction of other German immigrants, both Milwaukee and St. Louis became the major centers of the lager industry, with Pabst, Schlitz, Miller, and Blatz becoming giants in the former and Anheuser and Busch in the latter. Companies that came into being outside of these centers around this time were Hamm in St. Paul, Heileman in La Crosse, and Stroh in Detroit. Almost overnight, the Pilsner-style beers edged out the darker and richer beers that had first reached America with the English colonists.

More innovations came along to improve them. The development by Copenhagen's Carlsberg brewery of an absolutely pure brewer's yeast – which would end brewing failures – occurred in 1883. With the turn-of-the-century development of airtight kegs and carbonation, America's beers became bubbly, Pilsner-style beers, and in 1935, the Krueger company of New Jersey introduced the first canned beer (Trager 1995).

The Anheuser-Busch company may be said to embody the story of American beer. Formed by German immigrants Eberhard Anheuser and Adolphus Busch in the mid-nineteenth century, it capitalized on improved transportation and aggressive marketing techniques to the extent that, by 1901, the company was producing more than 1 million barrels of beer annually and had become the first to mass-market bottled beer. To provide the freshest beer available, Anheuser-Busch formed its own refrigerated railcar company and was one of the first brewers to employ pasteurization techniques (Smith 1995: 114–15). The company survived Prohibition by producing "near-beer" and malt for home brewers, maintained and even updated its brewing equipment, and emerged aggressively from those difficult times to become the giant it is today.

One reason for the popularity of lager beer in the United States is that hot summer days seem to call for ice-cold beverages, and the heavier beers did (and do) not lend themselves well to chilling. Presumably this explains why the British and many other Europeans drink beer that is at room temperature or perhaps cool but not cold. Another reason, however, is taste. Many believe that chilling removes taste, and in fact, most non-Americans do not particularly care for American beer, which they find to be uniformly bland in taste and lacking in character. European Pilsners contain less in the way of chemicals and generally more in the way of alcohol than American beers. On the other hand, Americans in general – and not just

those in the United States – enjoy cold lagers; in fact, Mexico and Brazil are among the world's top beer-producing countries. Mexican breweries in particular make a wide selection of light-tasting lagers, with perhaps Corona, along with Dos Equis (a dark Pilsner), among the best known (Pepper 1996).

It was Americans who introduced Foster's beer to Australia in 1888 and did it "American style" with refrigeration, bottom fermentation, and bottling – in the process creating a product that became Australia's national drink. In Jamaica, by contrast, the famous Red Stripe continued under English influence to be a dark, alelike brew until 1934, after which it finally was transformed into a light-tasting lager (Trager 1995).

Beer was brought to East Africa by the British in 1922 and is brewed there today mainly by East Africa Breweries in both Kenya and Uganda. Kenya Breweries, a subsidiary of East Africa Breweries, produces the Tusker lagers, some of the best-known beers in Africa (Pepper 1996: 135). Nigeria and South Africa, however, are the major beer producers on the African continent. Most of Nigeria's beer is brewed by Guinness and Heineken (both of which have major stakes in Nigeria), along with several indigenous breweries. South Africa's beer is produced by South African Breweries and is almost entirely in the lager style (Pepper 1996: 135).

Lager beer reached Asia in 1904, when German and British entrepreneurs established a Western-style brewery along the coast of northern China, producing a brand known as Tsingtao. The Dai Nippon Beer Company of Japan acquired the Tsingtao Brewery in 1916 and retained it until after World War II. In 1932, Japan continued the spread of lager in East Asia by constructing another brewery in Manchuria and introduced breweries into Korea as well. The Dai Nippon Beer Company, along with Kirin, dominated the Japanese beer market until broken up by American occupation authorities in the aftermath of World War II (Laker 1986: 60, 1987: 25–8).

Throughout the twentieth century, the trend has been toward ever larger commercial brewers. In the United States, for example, the number of large breweries has decreased from more than 200 to less than 50. Today, it is mostly lager beers that are consumed globally, and these are produced by big corporations for large markets – often international ones. Although European and American lagers tend to dominate the world market, one exception to this trend is Singha, a lager from Thailand, which is one of the few Asian beers that enjoy a wide degree of export to the West, especially as Thai restaurants continue to grow in popularity (Pepper 1996: 130).

The exceptions to the dominance of mass-produced, globally marketed beers are the products of some "microbreweries" (but not those connected with the giant brewing companies) that found a ready market in the 1980s and 1990s, when entrepreneurs began distributing innumerable beers that

were produced either in a "brew pub" or in a small brewery. Microbrewers offer a uniqueness not found in mass-produced beers, crafting their microbrews in a fashion reminiscent of earlier days, when beer production was confined to monasteries, village breweries, or to the home itself. Generally priced higher than their mass-produced counterparts but with a more refined appeal, microbrews have established a niche among the more affluent consumers in America and Europe.

Phillip A. Cantrell II

Bibliography

- Baron, Stanley Wade. 1962. *Brewed in America: A history of beer and ale in the United States*. Boston, Mass.
- Bower, Bruce. 1992. Vessel residue taps into early brewing. *Science News* 142: 310.
1994. Ancient site taps into soldiers' brew. *Science News* 146: 390.
- Clark, Peter. 1983. *The English alehouse: A social history, 1200–1830*. New York.
- Corran, H. S. 1975. *A history of brewing*. London.
- Dirar, Hamid A. 1993. *The indigenous foods of the Sudan: A study in African foods and nutrition*. Cambridge.
- Ghalioungui, P. 1979. Fermented beverages in antiquity. In *Fermented food beverages in nutrition*, ed. C. Gastineau, W. Darby, and T. Turner, 3–19. New York.
- Hawkins, K. H., and C. L. Pass. 1979. *The brewing industry: A study in industrial organization and public policy*. London.
- History with gusto. 1987. *Time*, April 6.
- Katz, Solomon H., and Mary M. Voigt. 1986. Bread and beer: The early use of cereals in the human diet. *Expedition* 28: 23–34.
- Kloss, C. A. 1959. *The art and science of brewing*. London.
- Laker, Joseph. 1986. Suds and substance: The Korean beer industry and Japan's influence on Korean industrialization. *Wheeling Jesuit College Annual*: 56–73.
1987. Green island beer. *Wheeling Jesuit College Annual*: 25–8.
- McGee, Harold. 1984. *On food and cooking*. New York.
- Morris, C. 1979. Maize beer in the economics, politics, and religion of the Inca empire. In *Fermented food beverages in nutrition*, ed. C. Gastineau, W. Darby, and T. Turner, 21–34. New York.
- Pepper, Barry. 1996. *The international book of beer: A guide to the world's most popular drink*. New York.
- Porter, John. 1975. *All about beer*. New York.
- Rorabaugh, W. J. 1979. *The alcoholic republic: An American tradition*. New York.
- Salem, F. W. 1972. *Beer: Its history and its economic value as a national beverage*. New York.
- Schivelbusch, Wolfgang. 1992. *Tastes of paradise: A social history of spices, stimulants, and intoxicants*, trans. David Jacobson. New York.
- Smith, Gregg. 1995. *Beer: A history of suds and civilization from Mesopotamia to microbreweries*. New York.
- Tannahill, Reay. 1988. *Food in history*. New York.
- Toussaint-Samat, Maguelonne. 1992. *History of food*, trans. Anthea Bell. Cambridge, Mass.
- Trager, James. 1995. *The food chronology*. New York.

III.2 Breast Milk and Artificial Infant Feeding

The importance of maternal breast feeding is considerable for infant survival until weaning and beyond. In addition to nutrition, it also provides many of the mother's immunological defenses to a baby whose own defenses are weak at birth. In the past, however, the protective (as well as the nutritive) qualities of maternal milk were unknown, and wet-nursing practices were common for social, cultural, and economic reasons. In the West, maternal breast feeding in the first year of life was most common in rural areas. But animal milks and solid food were given to babies everywhere when breast feeding was not possible, practical, or convenient, or when it was considered insufficient.

Maternal Breast Feeding

During the first year of life, mortality should be independent of the purely nutritional factor. It has been demonstrated that even undernourished mothers, subsisting on around 1,500 calories a day, are able to nourish their babies adequately through breast feeding (Livi-Bacci 1991). Nonetheless, in the past, infant mortality figures (which were usually high) varied enormously with maternal feeding practices. Babies not breast-fed are easily prone to gastrointestinal infections in summer and to respiratory infections in winter. Thus, environmental circumstances and child care were secondary factors of survival when compared with the manner in which infants were nourished.

Mother's milk is in itself complete nourishment for a new infant. However, compared to the milk of a new mother (which comes in response to the direct demand of the suckling child), the milk of a wet nurse has neither the "age" nor the exact composition required for neonates, because usually the nurse has given birth to a child of her own some months earlier. Nevertheless, female milk is preferable to animal milk. Our ancestors were perfectly aware of the age of the milk - its color and consistency - when choosing a wet nurse. Yet before the 1750s, they did not fully appreciate the qualities of human milk and the superiority of a mother's milk for her own child.

The Qualities of Human Milk

The constituents of human milk - water, protein, fat, carbohydrates, minerals, vitamins, lactose, casein, and so forth - have all been examined, and comparisons made with cow's milk and other animal milks (Paul and Palmer 1972; Davidson et al. 1975; Souci, Fachmann, and Kraut 1981). Such studies have demonstrated the advantages of human milk, and they show why cow's milk cannot safely be given to very young children unless it has been processed. This procedure

involves dilution with water, to reduce the excessive protein concentration, and also the addition of sugar. Moreover, before pasteurization, cow's milk presented other dangers. Since the milk was not boiled, and bottles and other instruments were not sterilized, artificial feeding could easily produce fatal infections. With direct sucking, however, such consequences were avoided.

The Wasted Colostrum

In European societies, there was a common and widespread belief that a mother's first milk, the colostrum, was not good for the newborn baby, who had to get rid of his or her meconium. Thus, another lactating woman frequently suckled the child for the first four or five days. Unfortunately, delayed lactation for new mothers could result in milk fever, painful milk retention, and unprepared nipples that could discourage her from ever breast-feeding (Fauve-Chamoux 1983).

Nor was such a procedure good for the baby. Although people of many cultures - from ancient Greeks to Mexican Indians (Soranus of Ephesus 1956; Kay 1981) - have shared the belief that the colostrum may be dangerous (indigestible, heavy, and corrupt), it is, although lightly purgative, actually well assimilated and very protective in the antibodies it delivers.

Yet until the middle of the eighteenth century, or even later, medical literature condemned the colostrum. Samuel Tissot, for example, in his best-selling handbook, *Avis au peuple sur sa santé* (1782: 379), wrote:

The stomach and bowels of the child, when he comes to the world, are full of a thick and sticky substance called *meconium*. This substance must be evacuated before the child is fed with milk, otherwise it would corrupt the milk; it would turn very bitter and produce two kinds of severe problems the child would not survive.

This excrement may be evacuated (1) if no milk is given during the first 24 hours. (2) If water is given during this time, with a little sugar or honey in it; resulting in a dilution of the meconium, a better natural evacuation and perhaps vomiting. (3) To ascertain that all the substance has gone, it is recommended to give, every four or five hours, an ounce of syrup made of *composite chicory*, added with a little water. This diet is of great advantage; it should be common practice, as it is now here [in Switzerland] since the last years; this syrup is better than any others, particularly better than almond's oil.

If the baby is very weak, and if some food is requested during the first day of life, some biscuit, soaked in water, is usually given, or a light *panade* [bread soup].

Most children of preindustrial rural families were fortunate enough to have this treatment succeeded by maternal breast feeding. But for many less-fortu-

nate urban children, it was followed by a regimen of wet nursing and complementary foods.

Control of the Milk Supply

According to the Greek physician Soranus, active in Rome and Alexandria about A.D. 100, a nurse's milk should be regularly tested for consistency and quantity. In order to maintain a good milk supply, any nursing woman had to follow a regimen of diet, rest, and some exercise. Earlier, around A.D. 60, Pliny the Elder in his *Naturalis Historia*, had indicated that some foods helped increase one's milk supply. These included cabbages, fennel, *Agnus castus* seeds, anemone, and the boiled stalk of sow thistle (Pliny 1601; Soranus 1956; Knibiehler and Fouquet 1980).

In late-nineteenth-century literature, it was often recommended that a lactating woman eat good meals, with cereals, lentils, and chocolate. In addition, it was believed that drinking beer, cider, or even coffee would help maintain a "humectant" diet (Delahaye 1990). However, it is interesting to note that, before the 1830s, such literature was astonishingly silent about beverages. From antiquity on, treatises recommending maternal breast feeding seem oblivious to the importance of drinking liquids. Yet 87 percent of human milk is water (Table III.2.1), and so a lactating mother should drink extra liquid to feed the child (Table III.2.2).

Indeed, this may be one of the reasons that so many urban women of the past complained about not being able to produce sufficient milk: They simply did not drink enough liquids. In European urban areas, water pollution was so general that water might have been used for cooking soup but not for drinking, and unboiled water was dangerous to health.

In rural areas, by contrast, peasants knew the respective qualities of various water sources and frequently had easier access to fresh spring water than urban dwellers, although it is true that even in the country, unusual pollution occasionally resulted in serious epidemics. The generally greater access to clean water supplies, however, helped ensure that breast feeding not only remained a predominantly

Table III.2.2. *Daily quantities of milk a healthy child should ordinarily absorb during the first six months of life (averages)*

Age of the child	G/suck	Number of sucklings	Total/day
First days	70	7	420
First weeks	90	7	630
2d month	120	6	720
3d month	140	6	840
4th month	180	5	900
5th month	240	4	960
6th month	120	4	480
	+ extra food		+ 200 water

Sources: Personal observation of a contemporary nurse, France (1981) and Fauve-Chamoux (1983).

rural practice but was also a paying occupation of peasant women, because they were regarded as fit to produce human milk on demand for several years at a time. In addition to earning money, country nurses probably also knew that lengthy lactation might help them avoid another pregnancy.

Delaying Conception: Milk Versus Blood

Acting through a complex hormonal mechanism, lactation can inhibit ovulation under optimal circumstances. However, both ethnographic and demographic data indicate that the relationship between breast feeding and anovulation is not perfect (Corsini 1979). In addition, other data help to disguise the importance of sexual abstinence during the postpartum period. In fact, many historical and/or demographic studies are difficult to interpret because they are mute about breast-feeding habits. When a mother is the sole source of infant nourishment, anovulation is frequent, and many populations have depended on the contraceptive effect of breast feeding for the spacing of births. Yet despite many recent historical studies using family reconstitution methods, the exact role of breast feeding in causing postpartum amenorrhea in past Western societies has eluded quantitative elucidation.

The age of the nurse, her nutrition, body weight, and fatness are all unknown parameters to which sexual behavior should be added (Knodel and Van de Walle 1967; Chamoux 1973b; Corsini 1974; Fauve-Chamoux 1983). Hippocrates and Soranus shared the idea that pregnancy could not occur without menstruation, the resumption of which, they thought, was encouraged by sexual intercourse, with the result that this specific blood altered the milk. Thus, many authorities insisted on the necessity of sexual abstinence during breast feeding (Van de Walle and Van de Walle 1972), but it seems that the great majority of European couples never accepted this taboo.

Soon after birth, part of the mother's blood was said to be transferred to the breasts as milk. "After helping a new life in the maternal uterus, blood goes

Table III.2.1. *Typical analyses of milk from various species*

	Human	Cow	Goat	Ass
Water (g/100ml)	87	87.5	86.5	91
Protein (g/100ml)	1.5	3.5	3.7	1.9
Fat (g/100ml)	4.0	3.5	4.8	1
Carbohydrates (g/100ml)	6.8	5.0	4.5	6.0
Minerals (g/100ml)	0.2	0.7	0.8	0.4
Lactose (g/100ml)	7	4.5	4.1	6.1
Energy (kcal)	68	66	76	41

Sources: Davidson, Passmore, Brock, and Truswell (1975), p. 236; and Souci, Fachmann, and Kraut (1981), pp. 2-20.

up to breasts, following an admirable, natural and economic way, and becomes a sweet and familiar food" (Verdier-Heurtin 1804: 30). Before conception, the female's extra blood was thought to be evacuated every month. Thus, after birth, if the mother did not lactate, her milk might be dangerously "driven back" and could make her sick. If menstruation or pregnancy occurred while wet-nursing, the milk was considered "spoiled" (Chamoux 1973b; Sussman 1982; Fauve-Chamoux 1985; Fildes 1988); in consequence, the milk of a pregnant nurse had a very bad reputation (Kay 1981; Fildes 1988).

Medical literature of the eighteenth century, which is replete with discussions of maternal breast feeding versus mercenary nursing, defended the Hippocratic principle of respect for nature: The free circulation of "humors" should be encouraged, and any impediment to this natural principle might create severe physical and moral troubles. Also common in the literature were references to the dangers of physical and moral contamination: "A child would have been honest if he had suckled his own mother's milk; instead he was a villain, a libertine, a vicious man with a bad temper; from suckling a bad wet-nurse he got all her vices and defaults" (Dionis 1721: 57).

A lascivious or hysterical nurse was believed to transmit her character through her milk, along with various eventual diseases – especially those of a venereal nature. Despite this, however, menstruating wet nurses were numerous, and those who were employed did their best to hide their monthly visitations (as well as any pregnancies) so as to continue nursing.

The Length of the Nursing Period

According to Soranus, the nursing period in his era ranged from 18 to 24 months – until the child could chew. Russian children of the eighteenth century were weaned after some 10 to 15 months of nursing (Chamoux 1973b; Ransel 1988). In Brittany at the beginning of the twentieth century, the weaning limit was 18 months, and "[t]hen my mother definitely closed her shirt and corselet," related Pierre Jakez Hélias (1975). "Then I had to eat my bouillie alone." Another tradition indicated that a child could be weaned when he had grown twenty teeth. In France, children of the rich could be suckled by their wet nurses until age 20 or 22 months.

By contrast, at the end of the ancien régime, the great majority of foundlings who survived were officially weaned when 12 months old. In the Rheims region, however, 54 percent of illegitimate babies were weaned when 12 months old, and 31 percent when 13 months old, for a very simple administrative reason: Wet nurses had a monthly salary, and thus the change took place on the first day of the month. But unfortunately, weaning concerned only the 30 percent of children who survived to secure access to a nurse's breast in the first place. Legitimate city chil-

dren sent to wet nurses in the country were not returned to Paris before 19 or 20 months of age, although they had been weaned some months before. Because treatises of the Enlightenment period did not recommend breast feeding exclusively (Hecquet 1708; Lerebourg 1767; L'Épinoï 1785), semisolid food was also a part of the infant's diet. Thus, some children could be weaned as early as 6 months or even less. The weaning schedule depended on a question of money: A dry nurse was cheaper than a wet nurse, and peasant women could officially sell their milk as soon as their own child reached 7 months (*Code des nourrices de Paris* [1715] cited in Delahaye 1990). It seems, however, that cheating was the norm in Paris as far as this regulation was concerned, and the age of the nurse's milk could be only a guess.

Mercenary Breast Feeding in Europe

Until the middle of the seventeenth century in Europe, the practice of putting a child out to nurse was reserved for the aristocracy. It was standard behavior to hand over an aristocrat's newborn child to a paid nurse, although the nurse was usually strictly supervised by the family. Much later, in the nineteenth century, bourgeois families utilized the system of an "on-the-spot" wet nurse who was given board and lodging and was, like her earlier counterpart, closely supervised. But in between these two periods – during the seventeenth and eighteenth centuries – newborn babies of those who could afford it were often sent to the country, and thus placed out of range of their mothers' control. We do not know how extensive this phenomenon was throughout Europe. In France, it was limited primarily to residents of the main towns (Chamoux 1973a), and in Renaissance Florence, there was a long tradition of urban parents sending infants out to nurse (*a balia*) soon after birth (Klapisch-Zuber 1985).

In preindustrial England, such a practice was never so widespread as in some other parts of Europe, but well-to-do urban families of merchants, lawyers, physicians, and clergymen – especially those residing in London – tended to employ country women a few miles away from the city (Fildes 1988). Similar situations existed in Warsaw, Hamburg, Stockholm, Vienna, and Madrid.

Parental Behavior and Social Norms

In the past, paternal feelings for children who had not yet reached the age of reason were frequently negative. Michel de Montaigne, for example, wrote in his *Essays* (1580), without mentioning his wife's opinion: "I cannot feel that passion necessary to kiss a barely-born child, with neither movement of his soul, nor a recognizable shape to its body, by which it might make itself pleasing to me. Nor have I been inclined to have them fed close by me." He continued that it was "a good school to send them [the small children] to

the country, to bring them up in the lowest and commonest way to live under popular and natural law.”

Basically, Montaigne’s “natural” educational model did not differ very much from that of Jean-Jacques Rousseau. Following the fame he had acquired from the articulation of his philosophy of child care in *Émile* (1762), in 1789 Rousseau wrote in his *Confessions* (1959): “I have neglected my duties as a father, but the desire to do them harm did not enter my heart, and my fatherly entrails could hardly cry out for children who have never been seen.” The latter remark was a reference to a wet-nursing arrangement that distanced the mother from her infant. As an earlier observer had lamented:

I am quite at a loss to account for the general practice of sending infants out of doors, to be suckled, or dry-nursed by another woman, who has not so much understanding, nor can have so much affection for it, as the parents; and how it comes to pass, that the people of good sense and easy circumstances will not give themselves the pains to watch over the health and welfare of their children (Cadogan 1748: 24).

Clearly, a bond between nurse and child frequently replaced that between mother and child, and many medical writers followed Soranus of Ephesus – the most famous gynecological and obstetrical writer of antiquity – who thought that a good wet nurse helped in protecting the mother’s health and youth. In the time of Soranus, however, the nurse was present in the family, and her behavior was under strict supervision. In the eighteenth century, this was not the case, and even a mother’s quest for elegance and parental desires for tranquillity cannot fully account for putting children out to nurse. Nonetheless, this was a widespread social phenomenon, which resulted in half the babies in Paris being sent outside the capital, no matter what their sex or birth order. The explanation seems to be that it had become common middle-class behavior.

Urban Females and Breast Feeding

Refusing the maternal duty of breast feeding, however, also extended to the lower classes, especially in preindustrial European towns. A French writer in Lyon in the 1860s explained:

Shopkeepers’ wives have, in general, as much importance in their business as their husbands, they cannot nurse for themselves; and besides, they usually live far from their shops, so they have to resort to the wet nursing bureaux. It is the same with female silk workers, who earn almost as much as their husbands and have an interest in putting their children out to nurse in the country (Sussman 1982: 104).

A century before, when women were working hard for the *manufactures*, a police officer in Lyon observed that “[t]heir situation, work, condition do

not allow the women any time, any freedom, to feed their own children” (Prost de Royer 1778).

There may, however, have been other reasons, not related to the work of the mothers, for sending children away. To send the newborn child to fresh air, in the belief that the bad urban air spoiled maternal milk, was doubtless one reason. Another may have been the perception that the diet of urban females did not contain enough in the way of appropriate beverages and good-quality proteins. Indeed, in France, consumption of meat in urban areas was very low at the end of the ancien régime (Livi-Bacci 1991). Thus, poor women of the countryside may even have had a better diet than their urban counterparts. Nonetheless, we are forced to wonder again what meaning “motherhood” held for those couples who put their children out to unsupervised country nurses, or what it meant for those mothers, often alone, who abandoned their babies to charity.

Love and Premodern Motherhood

It is doubtful that people of the past did not realize that putting a child out to nurse was risky, and that abandoning a child was to condemn it to death. This makes it difficult to explain how, in those days of enlightened philosophers, the widespread practice of abandonment of newborns (which in pre-Pasteur times usually amounted to infanticide) could coexist with a clear renewal of interest in children (Aries 1962; Shorter 1975; Badinter 1981; Bardet, Corsini, Perez Moreda, et al. 1991).

Indeed, it is morbidly fascinating that Rousseau, the herald of individual morality, seems to have had no difficulty adopting the behavior of contemporary Parisians with regard to bastard children: He abandoned his own children to the Paris foundling hospital. His reasons for this choice are very clear (Rousseau 1959), but the mentality that inspired them remains incomprehensible to us. Rousseau did not belong to any of the social groups regularly acquainted with child abandonment, such as domestic servants, urban workers, or impoverished day laborers, but his children were all bastards, and raising them was apparently not among his priorities. The same held for Thérèse Levasseur, his companion, who seemingly did not fight to keep her children, nor did she try to suckle her succession of babies, boys or girls. In Europe, as many boys were abandoned as girls (Bardet et al. 1991).

An abundant eighteenth-century literature encouraging maternal breast feeding stirred little interest in French, Italian, Russian, Polish, or English urban societies (Hecquet 1708; Lerebourg 1767; L’Epinoy 1785). It was the case that some grand ladies (who could afford to) breast-fed their children and then boasted about it as a way of returning to nature. Similarly, it was fashionable to have one’s portrait painted when breast-feeding, like a Madonna (Fauve-Chamoux 1985). All of this was a way to affirm one’s modernity and to symbolize an advanced feminism, but it was

also the case that men frequently did not wish their wives to breast-feed. Thus, often, aristocratic women who breast-fed their infants only did so during the day; at night, maids or nurses took care of the babies. The day for the child, the night for the husband; the sum was a compromise with God.

Serving the Husband before the Child

Since polygamy is not tolerated in most Christian societies, theologians and confessors have had great difficulty in establishing a hierarchy of sins: A woman was supposed to do everything possible to avoid a too-ardent husband, and, of course, also to avoid adultery. Yet neither could she refuse her wifely duty for too long, even if her aim was to preserve her milk, which helps to explain why confessors finally justified the practice of putting children out to nurse as being the lesser of two evils: "If the husband is in danger of lacking restraint, the woman must, if she is able, put her child out to nurse in order to tend to her husband's infirmity," said the *Abrégé du dictionnaire des cas de conscience* by M. Pontas (1764).

Certainly such a practice preserved the peace of the household by ensuring the father permanent access to his wife's bed. This aristocratic habit was popularized in France and reveals the importance placed by the Catholic Church on conjugal relations. The Church favored the relationship between the parents and relegated the child's interest to second place (Aries 1962; Noonan 1966; Flandrin 1973).

Religious attachment also influenced attitudes with respect to childbearing and nursing. Protestant women generally bore fewer children than their Catholic counterparts and nursed them more often themselves. Infant mortality was, thus, much higher in Catholic than in Protestant families during the seventeenth century (Knodel and Van de Walle 1967). But despite dechristianization and an enormous literature in favor of maternal care, there was no widespread return to maternal feeding at the end of the eighteenth century, and France remained a cornerstone of the nursing industry. Families characteristically had fewer children, but the practices of wet nursing and artificial feeding continued, and until World War I, the custom of putting infants out to nurse remained widespread and infant mortality remained high (Rollet-Echalier 1990). In short, the status of infants did not improve greatly in France (or in western Europe, for that matter) before the first decades of the twentieth century.

Differential Infant Mortality

The consequences of this apparent unconcern for the welfare of the young were of considerable importance. Putting babies out to nurse raised the infant mortality level and, in the absence of contraceptive practice, accelerated the rate of successive pregnancies. France's infant mortality, much higher than that of England, Sweden, or Denmark, was clearly a function of less breast feeding and poor child care (Table III.2.4).

Table III.2.3. *Number of babies, aged less than 1, abandoned in Paris Foundling Hospital, 1773-7, with infant mortality for each group, according to their origin*

	From Hôtel-Dieu	From Paris and the suburbs	From the provinces	Total
Babies abandoned	6,523	14,552	9,039	30,114
Died under 31 days	5,395	10,650	5,854	21,899
Died within 12 months	5,892	12,185	7,180	25,257
Died under 31 days	82.7%	73.1%	64.7%	72.7%
Died within 12 months	90.3%	83.7%	79.4%	83.8%

Sources: Dr. Tenon, Bibliothèque Nationale, Paris, NAF 22746, cited in Chamoux (1973b).

Among Parisian craftsmen who kept their children in the eighteenth century, one baby out of every four died before his or her first birthday. But in Lyon in the 1780s, the "Society for Helping Poor Mothers" paid women 9 pounds a month if they breast-fed their own children during the first year of life. The infant mortality level soon fell to 1 out of 6 babies; the drop was considered a large success. By contrast, however, when infants were systematically put out to nurse, the infant mortality rate at least doubled. Thus, 40 percent of legitimate babies in Rouen died if they were put out to a wet nurse. But when their mothers nursed them, infant mortality dropped to 20 percent, showing the vital importance of maternal breast feeding (Bardet 1973). In Lyon, only one-third came back alive from wet nursing, and in Paris, only between one-half and one-third of legitimate children survived mercenary nursing (Prost de Royer 1778).

From the Refusal of Breast Feeding to Contraception

Rejection of the newborn was accompanied in France by a progressive diminution of family size. In Rheims, at the end of the seventeenth century, women who married before age 20 bore an average of 11 or 12 children. But in the late eighteenth century, on the eve of the French Revolution, women were having only 6 children (Fauve-Chamoux 1985). The conjugal practices of urban elite groups had undoubtedly shown the way with regard to contraception. As early as the 1670s, the leading families of large cities like Rheims or Rouen (Bardet 1973; Fauve-Chamoux 1985) were successfully limiting the number of their children, and social differences in this regard were strongly marked.

During the eighteenth century, behavior that would later be called "Malthusian" took root in the middle class. This was a "family-planning" outlook that

implied deep changes in attitude toward the Christian religion, and toward love and procreation. But, for the lower classes, little had changed. Poor mothers, frequently without husbands, often considered that they had no choice but to abandon their children, whether legitimate or not. And this custom continued to confront societies with moral as well as practical questions concerning the care and feeding of the young.

Artificial Infant Feeding

Feeding Foundlings

About one-third of the children abandoned in Paris during the eighteenth century came from the provinces and, thus, were at least a few days old and already sick and underfed. Many of these babies died before reaching the ancient hospital, the Hôtel-Dieu, which also served as a foundling home. Among other things, these deaths tended to skew mortality data and make it appear that, at the Hôtel-Dieu, the mortality of children from the provinces was lower than that of children born and abandoned in Paris (Table III.2.5). In 1772, the king prohibited this transfer of infants to Paris, but the flow continued. The ban was renewed in 1779 but with no better results, and at the end of the *ancien régime*, between 6,000 and 7,000 newborn babies were being abandoned in the capital each year (Table III.2.5).

The mortality of these children was horrifying. More than 80 percent died, with 7 out of 10 deaths occurring during the first month of life and resulting from lack of appropriate food and nursing (Table III.2.3). Three-quarters of those who died were estimated by the authorities to be of illegitimate birth. European urban hospitals of the time, like the Hôtel Dieu, were transformed into depositories for abandoned children.

Table III.2.4. *General infant and child mortality in four European countries during the second half of the eighteenth century*

France 1750-99	England 1750-99	Sweden 1750-99	Denmark 1780-1800
273	165	200	191
215	104	155	156

Source: Livi-Bacci (1991), p. 74.

Paradoxically, however, such institutional efforts may have encouraged the phenomenon of infant abandonment all over Europe (Bardet et al. 1991).

The scarcity of wet nurses for such children was general throughout Europe: Institutions paid low wages and the women they employed were poor, often unfit for proper breast feeding, and frequently without enough maternal milk. Therefore, French hospitals experimented with different methods of nourishing abandoned young children, trying cow's, goat's, and donkey's milk, and often combining the animal milks with water and barley or with "rice water." At the Hôtel-Dieu in Paris, and at similar institutions in Rheims and Milan (where abandoned children were numerous), feeding with animal milks led to the use of feeding bottles and special feeding horns, called *cornets*. These became part of a baby's equipment when leaving the foundling hospital for the wet nurse's home, and they provided a ready means of dry nursing and early weaning.

In Paris at the end of the eighteenth century, there were 208 cradles for foundlings that were permanently occupied. Each saw an ultra-rapid turnover, and each had a personal crystal bottle attached (Chamoux 1973b).

Animal Milk and Baby Food

Cow's milk, mixed with "barley water" in the Paris foundling hospital during the eighteenth century, probably contributed to the death of many children during the first weeks of life. "Wheat water" was also given as a separate drink. A common milk lightener was "lentil water," with a dilution as high as 50 percent for "healthy babies," according to the French physician Dr. J. Tenon,¹ who explained that this liquid was made available "in a porc bladder equipped with a teat or in a feeding-bottle when the children were crying" (Tenon manuscripts, Bibliothèque Nationale, Paris).

Milk from goats was widely known to be lighter, and therefore better, for infants than that of cows. A sixteenth-century observer near Bordeaux, for example, wrote of country women whose breasts did not contain enough milk to feed their infants and consequently called upon goats "to help them" (Montaigne 1580). In 1634, goats were employed by Paris's Hôtel Dieu to feed the foundlings, as had been done successfully in Italy in the previous century (Lallemand 1885).

Table III.2.5. *Number of children abandoned in Paris Foundling Hospital, 1773-7, according to their age and origin*

From Hôtel-Dieu	From Paris and the suburbs		From the provinces		
Newborn	Age less than 1	Age +1	Age less than 1	Age +1	Total
6,523	14,552	462	9,039	1,375	31,951

Source: Dr. Tenon, Bibliothèque Nationale, Paris, NAF 22746, cited in Chamoux (1973b).

Such nourishment was certainly good when the babies were directly suckling the animals. But unfortunately, bottles, with all their attendant germs, were often used as intermediaries, and if another liquid was added – for example, herb tea of *chiendent* (“couch grass”) – it made for a dangerous diet. Goats were easier to maintain in an urban institution than donkeys, whose milk had enjoyed a good reputation for its digestibility since Roman times (Table III.2.1). In Paris in the 1880s, donkeys were used for direct suckling at the stable of the *Hospice des Enfants Malades*. But they were troublesome because lactation could not be prolonged unless the animals fed their own foals, and the experiment did not last long. Direct animal suckling, however, was usually a practice in hospitals for syphilitic children, who, because of fear of contagion, were never breast-fed.

Wine was said to be a “tonic” when given in dilution and in small quantity. It was often employed when no milk was available, particularly when the child was being transported without any nurse. For example, it had the effect of keeping abandoned children quiet when they were taken by slow carriage to the Parisian foundling hospital from far away. But alcohol was also believed good for infants when they suffered from gastrointestinal complaints (Fildes 1988). Calvados, champagne, or other spirits may, on occasion, have been ritually added to feeding bottles to celebrate baptisms or like events.

European Bouillies

Solid food for infants consisted of a kind of porridge, called *bouillie* in French. This was a food usually reserved for a weaning diet and for older children. In much of France, wheat was preferred for the porridge (oats were considered a food for animals), but the use of corn was common in southern France after the end of the ancien régime. There, a porridge called *millias*, cooked in a pan, served as a main course for both adults and children. This dish was similar to Romanian *mamaliga* or Italian and Iberian *polenta* (Livi-Bacci 1991).

It is fortunate that such complementary foods were not strictly reserved for weaning or for mercenary nursing but rather were given, along with milk, to very young children in both European and colonial societies. This explains why *marasmus* – the result of a grossly deficient intake of both proteins and calories (Gift, Washbon, and Harrison 1972; Livi-Bacci 1991) – was not a common type of malnutrition in past Western societies. By contrast, many Occidental babies died of too much unaccustomed solid food and too little of their own mothers’ milk. One example of such an imbalance was a “biscuit soup” prepared every day in a London foundling hospital. The apothecary complained in 1759 that milk was often forgotten in the mixture: “The biscuit is directed to be boiled in water to a proper consistence and afterward

mixed with a proper proportion of milk, but it is now and has been given to the children without any milk at all” (Fildes 1988: 169).

During the 1770s, Tenon collected information from all over France and Europe about artificial baby food. He understood that any artificial food was usually a killer when the very young baby was not or was no longer breast-fed. “Those foods,” he wrote, “given as supplements to breast milk, as thick *bouillies* made of ordinary milk, flour, yolk, and wine, sweetened with sugar, destroy the positive effects from the very little milk they happen to suckle; those foods are too thick and too heavy for a delicate baby’s stomach” (Chamoux 1973b: 415).

When mother’s milk was lacking, Alsatian and German women gave their own children a cream made from rice or bread. In seventeenth-century Brittany, rural mothers often fed their children with a finger, after first warming the bread soup in their own mouths (Lemieux 1985). Certainly, in the past, the easiest food to give a baby, other than milk, was wet and mashed bread.

Such a *panade* figured prominently in the weaning process, which came about when the mother could not, or no longer wanted to, breast-feed her own child. This was often the case when a woman was wet-nursing another child. In Salpêtrière Hospital in 1755, for example, children of employed wet nurses were served a portion of *bouillie* in the morning and another in the evening, and given soup at lunch time. A “portion” was half of a big spoonful, and 17 portions were prepared twice a day in a cooking pot called a *bonet*. Most of those children, who rarely received mother’s milk, also consumed a daily supplement of a pint of animal milk.

Tenon, who commented so extensively on eighteenth-century baby food, was, for his time, a rare defender of cow’s milk, apparently because he was influenced by positive reports of Russian customs. Muscovite children, he wrote, were treated roughly but seemed to become accustomed very early to animal milk and solid food. Indeed, from the very first day of life, Russian babies were fed with salted, mashed bread and cow’s milk. “The woman in charge puts the milk in her own mouth to get it warm and spits it in a feeding horn for the baby. When aged ten months, the child is given *cacha*, a sort of gruel from *sarrazin* (buckwheat) – cooked with milk, a much better *bouillie* than the one made in our countries.” Three other kinds of traditional Russian baby foods and pastes are mentioned by Tenon. One was made of mashed bread, butter, and beer, another of mashed bread with honey, and still another of mashed bread with premasticated meat. And when the babies cried, they were given, as early as one month of age, a light narcotic made from poppy seeds, commonly used for making traditional pastries (Tenon Manuscripts, Bibliothèque Nationale, Paris).

This diet, Tenon noted with admiration, “never results in *edema*,” which is a well-known symptom of malnutrition still observed in twentieth-century developing countries. Sometimes it is called *kwashiorkor* and is the result of a deficiency of protein even when caloric intake is relatively adequate (Gift et al. 1972). In light of Tenon’s observation, it seems that Russian children may have enjoyed more good-quality protein in their diet than any dry-nursed babies in western Europe, who were never given meat or fish but only cereals, and who often suffered from edema.

It turns out, however, that Tenon was misled by the French edition of a book concerning the new foundling hospitals of Russia established by Catherine the Great; he believed that Russian women did not usually breast-feed their children (Betzy 1775). Today we know that at the end of the eighteenth century, the feeding of foundlings was as much a problem in Moscow and St. Petersburg as it was elsewhere in Europe. Fully 14,000 babies were registered annually by the Moscow Foundling Hospital (Ransel 1988), and their mortality was estimated to be at least 60 percent, which was roughly comparable to French figures.

During the 1880s, wet nursing became a common female profession in Russia: “The majority of new mothers from our area,” wrote a Russian physician, “do not think merely of how it would be best to nurse their baby but how they might best draw advantage from their breast milk; this unfortunate phenomenon is expressed in two forms – either becoming a [private] wet nurse or by taking a foundling, and sometimes not one but several” (Ransel 1988). Such a situation had long been common in most of Europe, but it was one that was in the process of change for the better.

The Swan Song of the Wet Nurse

Many significant changes in baby care occurred during the nineteenth century. Among the middle classes in Europe, in Paris and Lyon as well as in London, Madrid, Vienna, and Moscow, it became expected that the wet nurse would reside in the home of the infant (Fay-Sallos 1980; Sussman 1982; Fildes 1988). In England, however, during early Victorian times, when an unmarried wet nurse left her own baby of six weeks to sell her milk in a London bourgeois family, she condemned him or her to death. Infants of the poor still rarely survived artificial feeding, and the absence of human milk and early weaning sacrificed most illegitimate, as well as many legitimate, children.

In Britain, the Infant Life Protection Bill was passed in 1872, but this failed to stop baby farming and wet nursing. In France, the Roussel Law (1874) had no more immediate effect, but it did oblige families to register all children who were nursed outside

the home, to provide the name and address of the wet nurse, and to indicate the fate of the children. This permitted the calculation of regional and national statistics on infant survival and the impact of feeding methods. In Marseille, during 1875, 50 percent of parents employed wet nurses. From the data available, it would seem that 20 percent of wet nurses lived in the homes of the infants and 10 percent lived close to the baby’s city. But 70 percent lived far away from the baby’s home, sometimes more than 100 miles.

In the 1890s, the importance of pasteurization was understood, and artificial feeding at last became less dangerous. Nevertheless, some baby bottles were impossible to clean properly, especially those with a long tube attached to them.

Conclusion

Information from the World Fertility Survey (*Population Reports* 1986), as well as from recent studies conducted in developing countries where diarrheal diseases are a major problem for infants, shows how important breast feeding can be to infant survival (Carballo and Marin-Lira 1984). In most of western Europe, North America, and Japan, the potential impact of low breast-feeding rates has been attenuated by improved water supply and sanitary systems. But such major changes occurred less than a century ago with the advent of pasteurization techniques in conjunction with cultural, economic, and social changes. In fact, the so-called 3 Ps (Pasteur, Politics, and Progress) worked together to make artificial feeding safer. Therefore, little effort was exerted from the 1940s to the early 1970s to promote breast feeding, and by 1973, the prevalence of maternal breast feeding had fallen to 25 percent in the United States, with a mean duration of three months or less.

However, in the 1980s and 1990s, a return to breast feeding was noticeable in many industrialized countries, and especially among the better-educated mothers (La Leche International 1981). They became aware both of the immunological and antimicrobial properties of their maternal milk and of the “bifidus factor” concerning the intestinal flora of their babies. In addition, mothers understood that breast feeding benefits their infants with less exposure to contaminants, and finally, they realized that their own milk has superior nutritional properties to animal milks.

In the past, mothers had no such knowledge, however, and if babies died in innocence and purity they were supposedly blessed in heaven. Thus, putting children out to nurse remained a common practice during the nineteenth century throughout the Western world and particularly in France. In addition, before the systematic pasteurization of milk and the provision of sterilized bottles, artificial feeding frequently sent babies, especially those abandoned in European

institutions, to their graves. Maternal breast feeding remains, as it was in the past, a matter of culture, family norms, and religion.

Antoinette Fauve-Chamoux

Note

1. Dr. J. Tenon's manuscripts, from which I quote, may be found in the collections of the Bibliothèque Nationale, Paris. His project during the 1770s was to reform the French hospital system. His most important opinions about artificial infant feeding habits in late eighteenth-century Europe have been discussed by Chamoux (1973b).

Bibliography

- Aries, Philippe. 1962. *Centuries of childhood: A social history of family life*. New York.
- Badinter, Elisabeth. 1981. *The myth of motherhood. An historical view of the maternal instinct*, trans. R. DeGaris. London.
- Bardet, Jean-Pierre. 1973. Enfants abandonnés et enfants assistés. *Sur la population française du XVIIe et XVIIIe siècles*, 19–48. Paris.
- Bardet, J.-P., C. Corsini, V. Perez Moreda, et al. 1991. *Enfance abandonnée et société en Europe, XIVe-XXe siècle*. Rome.
- Betzy, M. 1775. *Le plan et les statuts des établissements ordonnés par Catherine II*. Amsterdam.
- Cadogan, William. 1748. *An essay upon nursing and the management of children*. London.
- Carballo, Manuel, and Angelica Marin-Lira. 1984. *Patterns of breast-feeding, child health and fertility*. Geneva.
- Chamoux, Antoinette. 1973a. L'enfance abandonnée à Reims à la fin du XVIIIe siècle. *Annales de Démographie Historique* 1973: 263–301.
- 1973b. L'allaitement artificiel. *Annales de Démographie Historique* 1973: 410–18.
- Corsini, Carlo. 1974. La fécondité naturelle de la femme mariée: Le cas des nourrices. *Genus* 30: 243–59.
1979. Is the fertility-reducing effect of lactation really substantial? In *Natural fertility*, ed. Henri Leridon and Jane Menken, 195–215. Liège.
- Davidson, Stanley, R. Passmore, J. F. Brock, and A. S. Truswell. 1975. *Human nutrition and dietetics*. Edinburgh.
- Delahaye, Marie-Claude. 1990. *Tétons et tétines. Histoire de l'allaitement*. Paris.
- Dionis, P. 1721. *Traité général des accouchements*. . . . Paris.
- Fauve-Chamoux, Antoinette. 1983. La femme devant l'allaitement. *Annales de Démographie Historique* 1983: 7–22.
1985. Innovation et comportement parental en milieu urbain XVe–XIXe siècles. *Annales (E.S.C.)* 5: 1023–39.
- Fay-Sallos, Fanny. 1980. *Les nourrices à Paris au XIXe siècle*. Paris.
- Fildes, Valerie. 1988. *Wet nursing. A history from antiquity to the present*. Oxford.
- Flandrin, Jean-Louis. 1973. L'attitude à l'égard du petit enfant et les conduites sexuelles dans la civilisation occidentale. *Annales de Démographie Historique* 1973: 143–210.
- Fuchs, Rachel G. 1984. *Abandoned children. Foundlings and child welfare in nineteenth-century France*. Albany, N.Y.
- Gift, Helen H., Marjorie B. Washbon, and Gail G. Harrison. 1972. *Nutrition, behavior and change*. Englewood Cliffs, NJ.
- Hecquet, Philippe. 1708. *De l'obligation aux mères de nourrir leurs enfants*. Paris.
- Hélias, Pierre Jakez. 1975. *Le cheval d'orgueil. Mémoires d'un breton du pays bigouden*. Paris.
- Kay, Margarita A. 1981. *Anthropology of human birth*. Philadelphia, Pa.
- Klapisch-Zuber, Christiane. 1985. Blood parents and milk parents: Wet nursing in Florence, 1300–1530. In *Women, family, and ritual Renaissance Italy*, trans. L. G. Cochrane, 132–64. Chicago.
- Knibiehler, Yvonne, and Catherine Fouquet. 1980. *L'histoire des mères du moyen-âge à nos jours*. Paris.
- Knodel, J., and E. Van de Walle. 1967. Breast-feeding, fertility and infant mortality. *Population Studies* 21: 109–31.
- La Leche International, ed. 1981. *The womanly art of breast-feeding*. New York.
- Lallemant, Leon. 1885. *Histoire des enfants abandonnés et délaissés: Étude sur la protection de l'enfance aux diverses époques de la civilisation*. Paris.
- Lemieux, Denise. 1985. *Les petits innocents. L'enfance en Nouvelle-France*. Quebec.
- L'Épino, Roze de. 1785. *Avis aux mères qui veulent allaiter*. Paris.
- Lerebourg, M. 1767. *Avis aux mères qui veulent nourrir leurs enfants*. Utrecht.
- Livi-Bacci, Massimo. 1991. *Population and nutrition. An essay on European demographic history*. Cambridge.
- Montaigne, Michel de. 1580. *Les essais*. Paris.
- Noonan, John T., Jr. 1966. *Contraception: A history of its treatment by the Catholic theologians and canonists*. Cambridge, Mass.
- Paré, Ambroise. 1573. *De la génération*. Paris.
- Paul, Pauline C., and Helen H. Palmer. 1972. *Food theory and applications*. New York.
- Pliny. 1601. *History of the world. Commonly called the natural history of Plinius secundus*, trans. P. Holland. London.
- Pontas, M. 1764. *Abrégé du dictionnaire des cas de conscience*. Paris.
- Population Reports*. 1986. Series M. no. 8 (October): 1–32.
- Prost de Royer, Antoine-François. 1778. *Mémoire sur la conservation des enfants*. Lyon.
- Ransel, David L. 1988. *Mothers of misery: Child abandonment in Russia*. Princeton, N.J.
- Rollet-Echalier, Catherine. 1990. *La politique à l'égard de la petite enfance sous la IIIe République*. Paris.
- Rousseau, Jean-Jacques. [1789] 1959. *Les confessions*. In *Oeuvres complètes*. Paris.
- Shorter, Edward. 1975. *The making of the modern family*. New York.
- Soranus of Ephesus. 1956. *Gynecology*, trans. O. Temkin et al. Baltimore, Md.
- Souci, S. W., W. Fachmann, and H. Kraut. 1981. *Food composition and nutrition tables 1981/82*. Stuttgart.
- Sussman, G. D. 1982. *Selling mothers' milk. The wet-nursing business in France 1715–1914*. London.
- Tissot, Samuel. 1782. *Avis au peuple sur sa santé*. Paris.

Van de Walle, Étienne, and Francine Van de Walle. 1972. Allaitement, stérilité et contraception: Les opinions jusqu'au XIXe siècle. *Population* 4-5: 685-701.
 Verdier-Heurtin, J. F. 1804. *Discours sur l'allaitement*. Paris.

III.3 🐉 Cacao

Origins, Varieties, and Cultivation

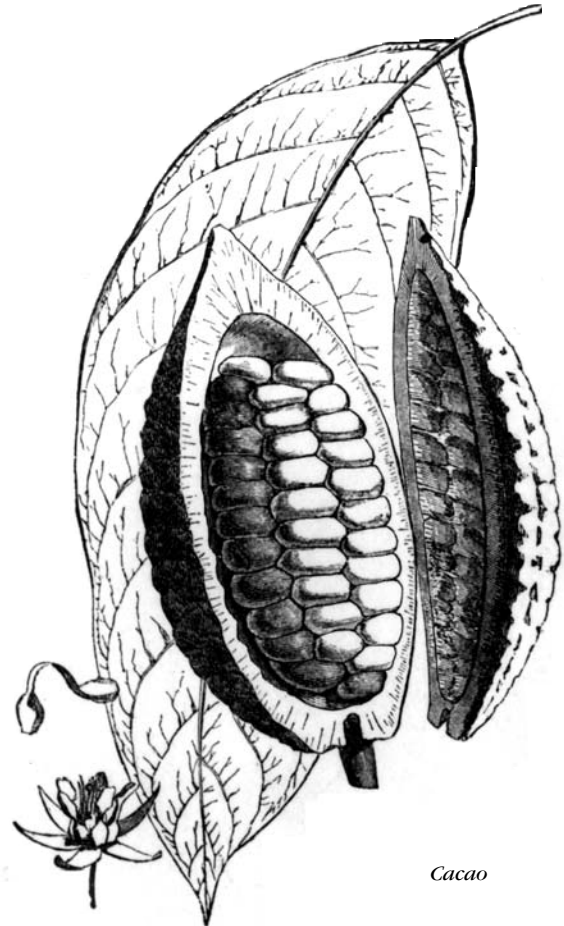
Cacao (*Theobroma cacao*), "the drink of the gods," and its main by-product, chocolate, are derived from the seeds of a fleshy pod, the fruit of the cacao tree. This tree is a tropical plant, certainly American and probably Amazonian in origin. In the Amazon region it is sometimes still found in its wild state, an understory plant usually well shaded by taller trees with dense foliage.

Today, cacao trees are sometimes grown in direct sunlight, thanks to modern fertilizers and hormonal treatments that help the trees produce a dense upper foliage. Most cacao trees, however, still require shade, and this is often provided by the simultaneous planting of shade trees and cacao saplings. Lemon trees, tall palms, and banana plants are employed, but more common is the aptly named *madre de cacao*, or "mother of the cacao" (*Gliricidia sepium*), another American native now found in all tropical areas. Various acacias and the coral tree have also been used to shade cacao plantations.

There is another species related to cacao, *Theobroma bicolor*, which is now a garden crop, although its pods were collected in the forest from ancient times until late in the Spanish-American colonial period. It does not produce true cacao and is known in Mesoamerica as *pataxte* or *patlaxtli*. Spaniards at first thought *patlaxtli* was harmful, but later they used the pods as a minor food and, in a few places, the tree to shade cacao trees.

There is some disagreement among experts as to cacao types. Many identify two varieties, others three, whereas a few claim that such distinctions evolved within historical times and were brought about by human selection and cultivation. For what it may be worth, most students of the subject agree that there are three types.

The *criollo* or "native" cacaos - that seem to have been grown only in Mesoamerica prior to the arrival of the Europeans - are considered to be the finest. All cacao trees demand considerable care, but *criollos* are the most difficult to cultivate. They are fragile, plagued by many pests, and relatively low in yield, and their cultivation today is limited to Mexico, Central America, Venezuela, and the Caribbean islands. *Criollo* cacao is often used to upgrade the quality of chocolate or



chocolate drinks, and connoisseurs place great stock in the percentage of *criollo* cacao in such confections.

Since the Spanish colonial period in America, *forastero* cacaos have been considered hardier and more prolific in yield, although of a lower quality than *criollo* cacaos. The meaning attached to the term *forastero* has varied, although it is always associated with wildness and alien origin. The plant may have spread from Brazil, but by the middle of the seventeenth century, it was thoroughly domesticated and widely cultivated in Ecuador and Venezuela. The Portuguese carried this variety from Brazil to São Tome and other offshore islands of West Africa, from whence it spread to the mainland. Today it is a leading crop in several African countries and in South-east Asia.

The third variety of cacao, according to several authorities, is a crossbreed between the two original types, and because it was first widely grown on the island of Trinidad, it has acquired the name *trinitario*. This variety is now dispersed and can be found as far from home as Indonesia and other parts of Asia.

Indeed, it may well be that the original Amazonian wild cacaos have become so geographically scattered and modified by human agency that they now appear to be highly distinctive varieties. Experts tend to identify cacaos and their various differences and desired qualities by regions, soils, and kinds of processing, so that Guayaquil cacaos, Brazilian cacaos, and varieties from the Ivory Coast or Malaysia, for example, are all characterized according to aroma, sweetness, the shape and size of the bean, oil or "butter" content, and other attributes.

Domesticated cacao trees are seldom allowed to grow to more than 20 feet in height. They are pruned and topped frequently to keep them accessible and to protect the cacao pods from sunlight and entangled branches. The pods develop from flowers that are fertilized naturally by midges, or artificially, and have the shape and rough size of an American football, although the size can vary greatly. The pod develops on the bark of the trunk or on large branches, not at the end of branches or twigs. The inside of the pod contains the beans or "nibs" and a white pulp, which is consumed by birds, animals, and humans. Cacao trees, although delicate, have a useful life of up to 50 years, but this can vary widely according to care, soil, region, and other circumstances. The trees develop slowly, usually from seedlings, and seldom yield any harvest until the third or fourth year. They are very sensitive to cold and drought and require year-round water from rain or irrigation.

Because cacao trees are too delicate to be climbed, harvesting must be done from the ground. Long poles topped by a cutting implement reach the higher fruit. After the pods are split open, the nibs are separated from the sticky pulp (by machine on modern plantations). The discarded pods are employed as fertilizer or animal fodder. The nibs are then fermented and dried. The workers heap them up in a shaded area and turn them over frequently for several days. Once fermentation is complete – the timing of which requires considerable expertise – the nibs are spread in the sun or dried artificially for a week or more, with workers again often raking and turning them.

The cacao nibs are then graded (nowadays by machine), poured into large sacks, inspected, and shipped to the major markets and consuming countries, where they are transformed and manufactured into hard chocolate, cacao powder, cacao butter, and other products.

A History of American Cacaos

The remote history of cacao is disputed; many deny its Amazonian origin and believe that in its "wild" state, the cacao tree grew in a number of other parts of the Americas, including perhaps the Upper Orinoco flood plain. Because the nibs have a short fertile life, there was difficulty in transporting them elsewhere to plant new seedlings, especially before modern transporta-

tion. Thus, gaps between areas of cultivation and different varieties came into existence.

In any event, it is in Mesoamerica that the first recorded histories of cacao are to be found. The plant and its nibs were part of the most ancient mythologies there. The Olmecs knew cacao, possibly before 1000 B.C., and the word itself, or one like it, may belong to a proto-Mixe-Zoquean language probably spoken by the Olmecs. Cacao was also part of the material and cultural lives of the Maya and their predecessors, as well as the various other societies (including that of the so-called Aztecs) that were destroyed by the Spanish invaders of Mesoamerica.

The Maya were probably the first people to write about cacao via their now-deciphered hieroglyphics. There may have been some link between cacao as a drink and the sacrifice of human blood. Certainly, elite burials and tombs often contained elaborate vessels with cacao drinks in them. Some scholars believe, in fact, that the consumption of cacao was a privilege of the nobility, but to others this practice seems unlikely (or at least not widespread), given the extent of its cultivation in various Mesoamerican tropical lowlands. Probably cacao was drunk (there is, so far, little evidence that it was eaten as a hard substance among Mesoamericans) by people of all classes. The drink was very much a part of public ceremony and ritual.

Cacao beans were an important item in Mesoamerican trade and tribute long before the European invasions. Cacao's climatic requirements, and the sophistication of state systems of trade and tribute, led to regional specialization – perhaps even monoculture – in a few places. In such areas as Guerrero, Colima, Veracruz, Tabasco, and the Gulf of Honduras, cacao was grown in plantation-like settings for export to distant centers. On his fourth voyage in 1502, Christopher Columbus in the Gulf of Honduras came upon a very large seagoing canoe that was transporting a varied cargo that included cacao beans. The most important growing areas, as far as present knowledge goes, were Soconusco (today the Pacific coast of Chiapas, Mexico) and the Pacific coastal plains and piedmonts of present-day Guatemala and El Salvador. From these plantations, large quantities of beans found their way to the highlands as trade and tribute. Soconusco, though distant from Tenochtitlan (now Mexico City), became a prized part of the Aztec conquests and paid tribute to Montezuma. Cacao beans were stored in great warehouses in the capital.

Money, therefore, "grew on trees." Cacao nibs were used as coinage throughout Mesoamerica from ancient times, and at least as far south as Costa Rica during the colonial period, especially when there were shortages of official coinage. Our knowledge of this coinage before the arrival of Europeans, however, is sketchy. In Nicaragua, a distant periphery of Mesoamerica, beans were widely used as petty coins and may have had standard equivalencies, evidence of a formal and official coinage. Additional evidence can be found in a

widely used serial system of measures that was based, not on weight, but upon the number of cacao nibs. There is also evidence of a thriving industry in counterfeit cacao beans, another sign of their monetary value. In some regions, however, the beans may have served only as substitute, specialized, or occasional coins.

Mesoamericans had many recipes for cacao drinks. Ground beans were mixed with hot or cold water and with maize, ground chillies, annatto, vanilla, and seeds, roots, and flowers of many kinds. Many of the dishes were soups, and the liquid chocolate, poured into or over many other dishes, was perhaps the ancestor of modern *mole* sauces. These peoples had a preference for chocolate drinks beaten to a froth, and to this end poured chocolate in a stream from one container to another. The upper classes and the *pochteca*, a kind of official merchant class, consumed large quantities of these soups and drinks at festivals and public banquets.

Hernán Cortés and his invading band noticed the many uses of cacao and captured stores of it in Tenochtitlan to use as money. They also saw the flourishing groves in such areas as Soconusco and Izalcos (in present-day El Salvador). In Soconusco, the harvests quickly attracted large numbers of mule trains and merchants. Some of the most lucrative *encomiendas* (grants of the labor and tribute of subject natives) in Central America were in Izalcos where cacao was generally the means of tribute-payment. But Spaniards usually did not try to seize legal ownership of cacao groves, perhaps because cultivation and care of the fragile, valuable trees was both a specialized business and hard work in a humid, tropical climate. Instead, they did their best to monopolize labor in the groves, not to mention the taxes cacao produced, and the cacao trade. Tribute to the crown and to private individuals continued to be paid in cacao long after the Spanish conquests, and large quantities were carried by sea and by land to central Mexico.

Gradually, the new colonial elites made changes in the "cacao economy" – not in the kinds of production or ownership, but rather in the intensity of cultivation and in patterns of distribution and consumption. *Encomenderos* (holders of *encomiendas*) and local royal officials forced native producers to intensify planting and harvesting, which may have been counterproductive for two reasons. First, the Native American population was already in severe decline because of the epidemiological invasion brought on by the arrival of Europeans and Africans. Overwork and other oppressions exacerbated this mortality, and replacements brought in from the highlands died just as quickly. By the late sixteenth century, many of the productive zones were suffering from severe labor shortages. Second, overplanting may also have been self-defeating. Shade trees were cut down, destroying the understory conditions needed by cacao trees. No doubt, labor shortages and changing agricultural conditions were related.

At the same time as production in Mesoamerica

declined, consumption rose. If cacao had once been a food limited to the upper classes, itself a doubtful proposition, this soon changed. Shortly after the conquest, it was apparent to some of the enterprising interlopers that cacao consumption was very large indeed; impressionistic accounts describe Indians quaffing massive, almost unbelievable, quantities of chocolate. Trade – often long-distance trade – in the crop became important, and demand was such that suppliers from faraway plantations could pay expensive freight costs and still show a profit.

As Central American production declined, prices in Mexico rose, and even more distant producers entered the market. The fine Central American *criollo* cacaos retained, and still retain, a market niche because of their reputed quality, but the Guayaquil basin of Ecuador and the coastal valleys of Venezuela west of Caracas, along the Tuy River, and south of Maracaibo, replaced Central America, Tabasco, and Guerrero as the largest producing areas. We know little about the origins of this new competition. In both areas, growers used the more hardy and productive *forastero* cacaos.

At first, Venezuelan cacao dominated. From around 1670, as the native populations disappeared, plantations were worked by African slaves. Exports from La Guaira, Maracaibo, and smaller ports were carried on a fleet of small ships to Veracruz, and from there by mule to Puebla and Mexico City. The Venezuelan elite became known as the "grandes cacaos," some of whom owned as many as a quarter of a million trees. In 1728, a clique of Basque entrepreneurs obtained an export monopoly called the Royal Guipuzcoa Company, or, colloquially, the Caracas Company. Between 1730 and 1754, exports nearly doubled to 3,800 tons per year, and a large part of the crop went no longer to Mexico but to Spain. Cacao also became part of a growing contraband trade, with smuggling by the Dutch especially important because of their colony at nearby Curaçao and because of a growing market for cacao in Amsterdam. In fact, Dutch traders dared to establish semipermanent trading depots in Venezuela well upstream on some of the rivers.

Guayaquil cacao plantations came into being at about the same time as those of Venezuela, and soon were shipping to Peru, Central America, and Mexico. Hardy varieties, plentiful water, good soils, and high yields contributed to the production of inexpensive cacaos, and Guayaquil prices, despite long trade routes, were able to undercut those of Central America and even Venezuela. Unlike the plantations in Venezuela, where African slaves were employed, Guayaquil cacao was worked by local laborers, often native migrants from the *sierra*.

Central American growers complained about the flood of inexpensive Guayaquil beans, and the Spanish crown, alarmed that a growing trade between Peru and Mexico would threaten its export trade in European goods to America, prohibited all South American

ports – including Guayaquil – from trading with Mexico. Again, contraband and fraud flourished. By 1700, Guayaquil's production rivaled or surpassed that of Venezuela. Later in the century, as free trade grew and faster ships became common, Guayaquil cacao began to reach Spain and other European markets, even though by the late eighteenth century, the region's plentiful exports were satisfying about three-quarters of the Mexican demand.

Amazonian cacao was slow to develop. This area of Brazil was neglected by Portugal and had few settlers. Lisbon was not a large market, and most of the trees, at least at first, were scattered in the forests. By the middle of the eighteenth century, however, Marañón cacao, worked mostly by African slaves, commanded a share of European markets.

Venezuelan cacao plantations spread to Trinidad, where the hybrid *trinitario* developed. Other parts of the circum-Caribbean, such as the coast of Costa Rica, Grenada, Martinique, and Saint-Domingue, also enjoyed ephemeral booms before production sank back, giving them only minor places in the Atlantic markets.

Changing Tastes and New Markets

The first Spaniards to encounter large stocks of cacao quickly realized (and exploited) its lucrative possibilities, especially in Mesoamerica. Their own taste for chocolate, however, developed more slowly. Native peoples sometimes added honey, but most of their recipes were bitter – even biting – to Spanish palates, especially when chilli peppers were added. English pirates, who were even less familiar with chocolate, burned or dumped the cargoes of cacao they captured. Sugar, introduced to the Americas by Spain and Portugal, helped to change minds, as did vanilla and other flavorings, although part of the story of Spanish and Portuguese acceptance of chocolate is also, no doubt, one of acculturation and acquired tastes. By the late sixteenth century, Europeans in the Americas were drinking sweet chocolate with great enthusiasm and had elaborated intricate and expensive containers, cups, and *molíneros*, or beating sticks, as frothy chocolate was still favored. So pervasive was the habit of drinking hot chocolate among Creole women that at least one bishop had to forbid it during Mass. An Englishman from Jamaica was amazed at the Spaniards' craving for chocolate drinks, and claimed that they seemed unable to live without them (Hans Sloane manuscript 11.410, folio 321, British Library).

Cacao beans probably reached Europe as early as the 1520s, shortly after the conquest of Mexico, but it was late in the century before regular cargoes were shipped from Veracruz. At first, chocolate was an expensive drink, largely confined to the Spanish court and noble houses. From Spain, it soon spread to Italy and – via a royal marriage – to France, where Ana of Austria, daughter of Philip III, married the young Louis XIII. The next marriage between French and Spanish

royalty was that of Louis XIV and the daughter of Philip IV of Spain, Maria Theresa, of whom it was said that her only passions were chocolate and His Majesty – probably in that order.

By the mid-seventeenth century, chocolate houses were common in Paris, although still patronized exclusively by the elite. They had also spread to London, where these establishments soon competed with coffee shops. During the life of Samuel Pepys, the English government official and well-known diarist, chocolate drinks passed from being a novelty to a regular luncheon beverage and a staple of public salons. The Dutch, who had captured Curaçao in 1634, began to send large cargoes of contraband Venezuelan cacao to Amsterdam, which became the great cacao mart of Europe. Ironically, by the late seventeenth century, even Spanish merchants had to buy there.

Chocolate, as a drink, is filling and quite nutritious, and so became popular among all classes in Catholic Europe, especially in the Mediterranean countries. But a problem arose: Was it to be considered a food or a beverage? Did the drinking of chocolate break a religious fast? In the 1660s, the Church decided that because wine – although nutritious – was considered a drink rather than a food, chocolate drinks would be treated the same way. Opinions differed as to chocolate's merits and failings, and these debates continue today. For reasons not yet understood, chocolate became a drink for women and coffee a drink for men in some parts of western Europe.

In the eighteenth century, chocolate was known throughout Europe, and further regional variations developed, as did cultural and fashionable attitudes toward it and its consumption. It was usually sold as a fairly solid paste or block, and already by the early part of the century, these commercial products contained vanilla, sugar, cloves, and other spices. About this time, some drinkers began to add milk or wine. Frothiness was still appreciated, and numerous recipes stipulated the amount of heating needed and the best way of whipping the chocolate to provide the ideal *frappé*.

The eating of solid chocolate, while still far behind the drink in popularity, also gained acceptance, often in the form of the so-called Spanish sticks. In France, small chocolate pastilles gained favor. Cacao butter, used since the sixteenth century as a cosmetic to improve the complexion, now became of some importance for elite cuisine.

In much of eighteenth-century Europe, chocolate remained an elite and rather expensive drink, sometimes considered a stimulant, or even an aphrodisiac (what exotic substance has not been so imagined?). In England, however, both coffee and chocolate houses lost their upper-class cachet in the second half of the century, and chocolate houses, in particular, came to be associated with raucous behavior, licentiousness, and marginal social groups. But even then, chocolate remained expensive; although American production had increased to supply the growing demand, it had

not increased enough to lower the price significantly. Moreover, processing techniques had failed to adapt. In the middle years of the eighteenth century, cacao was grown, harvested, elaborated, and sold much as it had been by Native Americans before Europeans arrived in Mexico.

Some innovations did appear. For example, Joseph S. Fry, who founded a company in England in 1728 (which lasted into the late twentieth century), created factory conditions for the mass production of chocolate products for the English market. Others in England and on the continent followed, but radical transformations in many other areas of the chocolate trade had to wait until the nineteenth and twentieth centuries.

The Industrial, Commercial, and Supply Revolutions

The Dutchman Coenrad Johannes Van Houten was the first in a line of inventors and entrepreneurs who modernized the manufacture of chocolate. He sought a better method for extracting the cacao “butter” or fats from the beans. He had started a factory in Amsterdam in 1815 and was unhappy with the traditional method of boiling the chocolate and then skimming off the fats. By 1828, he had invented and patented a dry press, which left much less butter in the end product. Van Houten then treated this “cocoa” powder with alkaline salts to make it more adaptable to various kinds of mixes, and very quickly, many of the older methods of manufacturing yielded to cocoa. Some twenty years later, an early form of “instant” chocolate had become widely accepted, and about midcentury, Joseph Fry and his sons seized upon this new cocoa powder and began to market the first commercial chocolate bars.

The Van Houten innovation was, historically and gastronomically speaking, just in time. By 1800, chocolate drinks had fallen well behind competing alkaloid beverages. In the West coffee dominated almost everywhere and has remained in the lead to the present day – except in England, where coffee first defeated chocolate and was then, in turn, defeated by tea. Chocolate was to survive and flourish, however, by going its separate way and leaving the field of adult drinks to its competitors. Except in a few of the Mediterranean countries and in parts of Latin America, where chocolate drinks retained popularity, chocolate now became cocoa, heavily sugared and drunk mostly by children and invalids, or chocolate bars and wrapped candy, devoured by all ages and classes.

The early great British rivals of the Fry Company were the Cadburys and Rowntrees. It was the firm of Cadbury, which quickly became the leading British confectioner, that first emphasized the connection between boxes of chocolate and romance. Chocolates and cut flowers became tokens of romantic love and

soon were the typical presents given to a woman. Sales of chocolate in boxes and specialty chocolates soared, and increase even more so today, especially on Valentine’s Day, at Christmas time, and at Easter – the season for chocolate eggs.

After the breakthroughs of the Dutch and English pioneers, Swiss inventors and businessmen took over. François Louis Cailler and Phillipe Suchard opened modern factories in the 1820s. Henri Nestlé, a chemist who first elaborated powdered milk, and Daniel Peter combined their talents to produce milk-chocolate bars, and Rodolphe Lindt, yet another Swiss, invented a process by which cacao was made smoother to the palate, more homogenized, and of finer aroma. Lindt’s “conching” process, which he began in 1879, radically changed the manufacture of chocolate, and his smooth chocolates soon replaced the rougher and grainier ones of the past. Swiss chocolate became the world standard for the chocolate bar, and such companies as Lindt and Nestlé are corporate giants today.

The country that was the scene for Henry Ford’s genius now saw the rise of the great entrepreneur of chocolate manufacture and sale. Milton Hershey, like Ford, saw the great possibilities of economies of scale and uniformity in producing and mass marketing for the millions. A confectioner from his youth, Hershey became convinced after a journey to Europe that he should concentrate on chocolate. He built a model workers’ town (named after himself) in Pennsylvania, on a much larger scale than those of Cadbury and Rowntree in England. To supply his huge chocolate factory, Hershey also transformed other landscapes: He boasted that the enormous dairy farms surrounding his town supplied 60,000 gallons of milk every working day. In Cuba, vast sugarcane estates were laid out east of Havana, complete with their own railroads and port terminals. (The Hershey empire in Cuba was later nationalized by the Castro government.) The Hershey Company was a model of modern manufacture and merchandising and soon controlled a giant share of the U.S. market.

In spite of mass manufacturers such as Cadbury and Hershey, producers of fine chocolates managed to retain a respectable and profitable share of the market, which has increased in recent years. High-quality chocolates, such as Godiva of Belgium and Valrhona of France, are known worldwide. Chocolate has entered the world of great cuisine, too. Profiteroles, chocolate cakes and pastries, mousses, and humble chocolate ice cream all have fanatical addicts.

Just as Hershey’s chocolate for the masses created dairy farms and sugar plantations, it also led to the expansion of cacao production. Cacao has become a world commodity, grown in all tropical regions, and traded and speculated upon in the stock and futures markets of London, New York, Paris, and other financial hubs.

Venezuelan exports, which had seemed so well placed geographically and logistically to supply

European markets during the Spanish colonial years, began to decline around 1800. The wars of independence there were enormously destructive, and Venezuelan producers found themselves less able to hold their own against both old and new competitors. Nonetheless, although coffee replaced cacao as the country's main export after 1830, Venezuela retained a respectable place as a supplier of cacao until about 1900.

It was the Guayaquil plantations of Ecuador, however, that satisfied the largest part of the world's needs in the nineteenth century. The Panama Canal opened, and Ecuador was fortunate (or unfortunate) to have an abundant, inexpensive labor supply from the highlands, where the failure of the textile industry in the face of British competitive imports had sent waves of unemployed workers to the coast. Cacao was promoted by early leaders of the newly independent republic, including its first president, Juan José Flores (1800-64), himself an owner of large plantations. There was a plateau in production - caused by civil war and yellow fever - between 1840 and 1870, but then output again accelerated, reaching more than 20,000 tons per year by 1900. In the first decades of the twentieth century, however, Ecuador's predominance in cacao production finally yielded to two major factors. Witchbroom disease destroyed many of the plantings in the 1920s, and the 1930s saw depressed world prices. As a consequence, new areas in other parts of the world began to compete.

The first threat to Latin American producers and exporters came from islands off the west coast of Africa, where cacao had been introduced to São Tome and Principe by the Portuguese. Development of the crop's products, however, suffered considerable delay because the islands continued to concentrate on coffee growing with slave labor until that industry suffered a crisis brought on by the abolition of slavery in 1875. Chocolate then took over, although working conditions on the cacao plantations were little better than those of slavery. By 1905, the two islands were the world's leading exporters of cacao beans.

This success aroused interest in the nearby mainland British colonies, although it was often the local people, rather than the colonial government, who developed the crop. By the 1920s, the Gold Coast (now Ghana) led the world in output, and Africa had passed America in cacao exports. The growth in demand during the first 40 years of this century was astonishing. World trade in the commodity went from about 100,000 tons in 1900 to 786,000 tons in 1939, a growth of nearly 800 percent! Nigeria and the Ivory Coast were now rivals of the Gold Coast in cacao production, and other colonies, such as Gabon and the Belgian Congo, were becoming exporters. By the 1990s, the Ivory Coast was the world's leading producer.

Nonetheless, West Africa was losing its dominant position. It produced three-quarters of the world's

cacao in 1960, but just over half by the 1990s. Latin American plantings began to revive somewhat, especially on new establishments around Bahia, but the latest rival areas were in Indonesia, and, above all, Malaysia, which by the early 1990s was the world's fourth leading producer of cacao, with annual exports of about 220,000 tons.

The world market in cacao remained impressive at about 2,500,000 tons per year in the 1990s, but that amount was much less than half of the total for coffee and represented just over 2 percent of the world trade in sugar. A constant problem with cacao has been a lack of price stability. Prices have fluctuated widely since World War II, and stabilization agreements have had little success. The United States is by far the leading consumer of cacao, followed by western Europe, Russia, and Japan. Per capita consumption in the United States and western Europe has doubled since 1945. The Swiss eat the most chocolate, followed by the British. The Norwegians and Austrians lead the world as drinkers of chocolate.

To what does chocolate owe its success? It is, for one thing, a well-rounded food, containing glucose, lipids, and proteins. It also provides significant quantities of important minerals, such as potassium and calcium. Of course, such additives as milk and sugar bring other qualities and calories. Cacao contains several stimulants, led by theobromine, then caffeine and serotonin.

Debates based on more impressionistic evidence have raged. Chocolate was long ago accused of being a cause of juvenile acne, but this now appears to be untrue. The argument that chocolate can elevate cholesterol levels has not yet been resolved. Is chocolate addictive? "Chocoholics" would certainly answer in the affirmative. So, we suppose, would the native peoples of sixteenth-century Mesoamerica. Recently, researchers reported that cacao consumption stimulates a mild, marijuanalike effect, reducing stress and encouraging a harmless euphoria. Debates and research continue, and meanwhile, the product's market range expands. Attempts to find substitutes enjoy little success, and the "food of the gods" continues to delight the millions.

Murdo J. MacLeod

Bibliography

- Alden, Dauril. 1976. *The significance of cacao production in the Amazon region during the late colonial period: An essay in comparative economic history*. Philadelphia, Pa.
- Assoumou, Jean. 1977. *L'économie du cacao*. Paris.
- Bergman, James L. 1969. The distribution of cacao cultivation in pre-Columbian America. *Annals of the Association of American Geographers* 59: 85-96.

- Chiriboga, Manuel. 1980. *Jornaleros y gran propietarios en 135 años de exportación cacaotera, 1790-1925*. Quito.
- Coady, Chantal. 1995. *The chocolate companion: A connoisseur's guide to the world's finest chocolates*. New York.
- Coe, Sophie D., and Michael D. Coe. 1996. *The true history of chocolate*. New York.
- Fuller, Linda K. 1994. *Chocolate fads, folklore and fantasies: 1000+ chunks of chocolate*. New York.
- Harwich, Nikita. 1992. *Histoire du chocolat*. Paris.
- Piñero, Eugenio. 1994. *The town of San Felipe and colonial cacao economies*. Philadelphia, Pa.
- Young, Allen M. 1994. *The chocolate tree. A natural history of cacao*. Washington, D.C.

III.4 ☛ Coffee

Coffee is a tree or bush that originated in Africa. It was first domesticated in Arabia, and massively consumed in Europe and North America. Later grown in Asia and Latin America, coffee, more than any other crop, has tied together the rich and the poor. Originally a luxury, coffee has become a necessity for consumer and producer alike and, in terms of value, is one of the leading internationally traded commodities today and probably has been the most important internationally traded agricultural product in history.

Coffee, however, has also been one of the most contradictory and controversial of crops. It has linked the religious and the secular, the archaic and the bourgeois, the proletarian and the intellectual, the enslaved and the free, the laborer and the dilettante. It has been accused of destroying societies, of perpetuating vice, and of undermining developing economies.

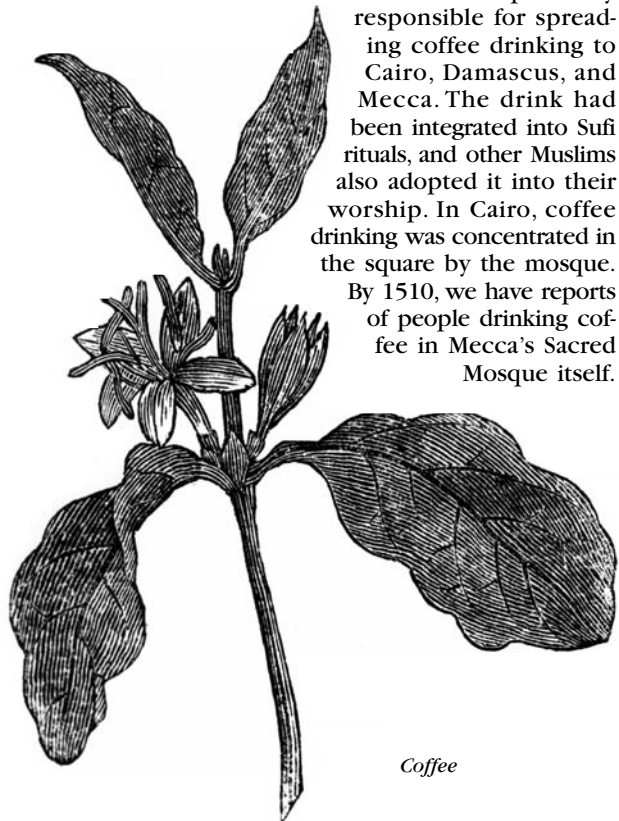
Origins

The first human consumption of coffee has been obscured by time. But legends abound, such as that of a ninth century A.D. Ethiopian goatherd who tasted the bitter berries that left his flock animated, or about Arab traders, and even Christian monks, who first recognized the virtues of coffee. *Coffea arabica* first appeared natively in Ethiopia, yet the berries went largely ignored before Arabs in Yemen used them to brew a drink. Although some Africans drank coffee made from fresh berries, others roasted it with melted butter, and in a few regions it was chewed without any preparation. However, no extensive local traditions of arabica berry usage developed (see Ukers 1948; Uribe 1954), and consequently, coffee became an exotic crop, growing far from its original home. Indeed, it was only in the twentieth century that African coffee production became substantial.

The mystical Shadhili Sufi, to the east of Ethiopia in Yemen, seem to have been among the first to embrace this wandering crop. They sought elixirs to produce visions granting access to the Godhead, and they also wished to remain awake during their long nighttime chanting rituals, in which they attempted to enter an ecstatic trance. Thus, from the beginning, coffee was employed as both a drug and a social drink. Early accounts, however, speak much more of its physical effects than its taste or efficacy in quenching thirst. A sheikh of the Sufi order who lived in the port town of Mocca in the late fourteenth or early fifteenth century may have been the first to devise a technique for roasting, grinding, and brewing coffee beans.

It was in the mountains of northern Yemen that the arabica was first domesticated, and for two and a half centuries Yemen held a virtual world monopoly on coffee production. But it is interesting to note that in Yemen, many preferred to chew the beans or brew a tea with the husk of *Coffea arabica*, rather than use the beans, and to this day Yemenis do not drink much coffee. They prefer tea or chewing a shrub, *kbat*, which not only fills coffee's social role but occupies much of the land suitable for coffee cultivation.

By 1500, the beverage had become widespread on the Arabian peninsula (Hattox 1985; Wenner 1991), with the Sufi probably responsible for spreading coffee drinking to Cairo, Damascus, and Mecca. The drink had been integrated into Sufi rituals, and other Muslims also adopted it into their worship. In Cairo, coffee drinking was concentrated in the square by the mosque. By 1510, we have reports of people drinking coffee in Mecca's Sacred Mosque itself.



The beverage became associated with Mohammed's birthday and was commonly drunk at night during the monthlong fast of Ramadan. Indeed, various legends ascribed coffee's origins to Mohammed, who, through the Archangel Gabriel, gave it to the world to replace the wine that Islam forbade. Certainly, Muslims were instrumental in spreading the beverage throughout the Islamic world as far as India and Indonesia, as the religious brought beans back from their pilgrimages to Mecca (Becker, Hoehfeld, and Kopp 1979).

Coffee was also intimately related to the growth of secular society. The Sufi were not full-time mystics, and as ordinary tradespeople during the day, they spread their taste for the drink to the business world. Its name probably comes from the Arabic *qahwah* – an epithet for “wine” – indicating a replacement of the forbidden alcoholic beverage. Indeed, although coffee grows wild in the region of Ethiopia known as Kaffa, the place is not the origin of its name, just as the coffee “bean” is not a bean but more like a cherry pit. Its name is a corruption of the Arabic for the arabica plant, *bunn*.

As already noted, the thick, dark, hot beverage became popular probably more because of its physical effects and the sociability it encouraged than because of its taste, although its reputed medicinal properties, said to cure mange, may have increased its appeal. But despite the fact that sugar had been grown in the Middle East for hundreds of years before coffee's arrival, it was not added to the brew. Neither was milk, which was blamed for causing leprosy when combined with *qahwah*. In fact, cardamom was the only spice added with any frequency, although in some disreputable quarters, opium and hashish apparently were also stirred in. The fact that coffee grounds must have boiling water poured over them to impart their flavor probably added to the drink's attractiveness. In an era before safe water supplies, boiling was the one sure method of insuring purity.

Probably the technology of coffee making – roasting and grinding the beans and boiling water – caused the beverage, more than almost any other commodity, to be associated with a site. The café was born in the Middle East, and in many cultures that subsequently adopted the beverage, the word for the site and the beverage became the same. R. Hattox (1985) suggests that early merchants, such as two Syrians who introduced coffee from Egypt to Istanbul in 1555, used the coffeehouse as a marketing device. To acquaint new consumers with the beverage, it had to be presented to them hot and correctly brewed. Consequently, these Syrian merchants and many others from the Levant who opened cafés found, in the process, an unoccupied niche in Middle Eastern society. Restaurants were almost unknown, and taverns were forbidden to Muslims. Hence, coffeehouses became one of the few secular public places in Muslim lands where men could congregate with other nonfamily members.

Cafés offered one of the only social possibilities for nightlife. They were an inexpensive place to offer friends entertainment and hospitality, playing an important role in the commodification of what previously had been largely private, domestic functions. Men read, listened to stories and music, watched puppet shows, played backgammon and chess, gambled, and sometimes solicited prostitutes. Such an ambience also led to conversations on forbidden subjects.

The cafés catered to a wide range of clients, sometimes mixing men of various classes. Although some cafés were little more than stalls in the marketplace, others were virtual pleasure palaces. Said the French traveler Philippe Dufour in 1685: “All the cafes of Damascus are beautiful – many fountains, nearby rivers, in tree-shaded spots with roses and other flowers” (Hattox 1985: 81). But because coffeehouses in Cairo, Istanbul, Damascus, and Algiers quickly became centers of political intrigue and fleshly vice (Carlier 1990), government officials soon reacted, and as early as 1511, they burned bags of coffee in the streets of Mecca.

Although some medical authorities claimed that the arabica's cold and dry humors created melancholy and, more seriously, transgressed the laws of Islam by intoxicating its users, these arguments never carried much weight. It was the politically subversive possibilities of the coffeehouse that most worried Middle Eastern rulers. Thus, the Turkish grand vizier was sufficiently concerned about the political effects of the 600 coffeehouses in Istanbul that he decreed that the punishment for operating a coffeehouse was cudgeling; for a second offense, the perpetrator was sewn into a leather bag and thrown into the Bosphorus.

Even these draconian measures proved ineffectual, however, and Turkish coffee became famous as the generic name for a certain thick brew in which the grinds were mixed directly into boiling water in small increments, often 10 to 12 times. The Middle East and Southeast Asia were the world's principal coffee drinking areas until roughly the middle of the eighteenth century. But ironically, today the Turks produce and consume virtually no coffee.

European Consumption

The increasing popularity of coffee in seventeenth-century Europe paralleled the emergence of commercial capitalism. The Middle Eastern bean, fostered by the ascetic Sufi to free themselves from worldly matters, evolved into a Western capitalist commodity. Coffee reached Europe via trade, diplomacy, war, and immigration. Venetian traders, with their command of Mediterranean commerce, introduced it into southern Europe. The Dutch, who superseded them in the Orient, were even more important in spreading the arabica.

At first, coffee was regarded in Europe as a medicinal drug that could cure sore eyes, dropsy, gout, and

scurvy. Its social role and prestige, however, were considerably enhanced in 1665 and 1666 by the arrival in France of emissaries of the Ottoman sultan who poured the exotic liquor for their aristocratic European guests during extravagant soirees (Leclant 1979; Heise 1987).

According to Austrian lore, coffee first became a drink Europeans found palatable when bags of it were left behind by the Ottoman Turks after their 1683 siege of Vienna failed to break the Austrians' spirit. These spoils of war not only uplifted Viennese spirits but also helped transform coffee from medicine into a leisure drink when the owner of the first Viennese coffeehouse, Georg Kolshitsky, thought to remove the sediment from Turkish coffee and add honey and milk. Elsewhere in Europe, Armenians, Greeks, Lebanese, and other Christian traders from the Levant spread knowledge of the beverage. Southern and central Europeans devised many of the most popular ways of brewing, roasting, and mixing coffee: Espresso, cappuccino, café au lait, and French and Vienna roasts are still familiar terms today.

Yet it was the northern Europeans who became the greatest consumers. In England, coffee was closely tied to the academic community from the beginning: The country's first coffeehouse seems to have been one opened in Oxford in 1637 by a Greek merchant. But soon London merchants were also imbibing the potion in coffeehouses, such as Jonathan's and Garraway's (which served for three-quarters of a century as England's main stock exchanges), the Virginia, the Baltic (which doubled as the mercantile shipping exchange), and Lloyd's Café (which became the world's largest insurance company). In addition to serving as commercial houses and office buildings, coffeehouses became "penny universities" that disseminated the latest news, reading libraries, and the first men's clubs.

Such social areas outside of the home and the court helped stimulate business but also outraged wives. They resented their husbands' addiction to the dark, noisy coffeehouses and the black and nauseous liquid that allegedly caused impotence. Their complaint, which was echoed in other corners of Europe, was in striking contrast to the mullahs' fears that coffee stimulated carnal desires. Charles II, concerned more with café patrons' political discussions than their familial responsibilities, tried unsuccessfully to close down the coffeehouses. It would take the rise of the East India Company, the Indian colonies, and high taxes on coffee to convert Britain to a tea-consuming country (Ellis 1956; Bramah 1972).

On the Continent, cafés symbolized and served the beneficiaries of capitalist prosperity who constituted the new leisure class, although debates continued to rage over the brew's medicinal properties. In the best scientific tradition, Sweden's Gustav III reputedly commuted the death sentences of twin brothers convicted of murder on the condition that one be given

only tea to drink and the other coffee. The tea drinker died first – at age 83 – and Sweden became the world's most ardent coffee-consuming nation, with its citizens drinking five cups per person per day by 1975. In 1992, most of the 10 leading coffee-consuming nations on a per capita basis were still in northern Europe: The 10 were (in order) Finland, Sweden, Denmark, Norway, Netherlands, Switzerland, Austria, Germany, France, and Belgium (United States Department of Agriculture 1993).

The eighteenth-century mercantilist policies of tax-hungry kings posed another kind of threat to coffee drinkers. Frederick the Great, for example, was less open-minded than King Gustav and less concerned with his subjects' health than with their political proclivities and the balance of trade. He consequently sought to prevent commoners from drinking the brew by making it a royal monopoly. Although he failed, the high import duties in the seventeenth and eighteenth centuries restricted consumption to the relatively affluent in major cities. The same was true in France and Austria (Heise 1987) and in Switzerland, between 1756 and 1822, there were five different decrees prohibiting coffee importation.

Cafés prospered in the capitals, however. Their great popularity in Paris served to distinguish the elite from their social inferiors, who bought their coffee in the marketplace. The coffeehouse denizens constituted an elite of achievement, a bourgeois elite. Coffee's great virtue, in contradistinction to alcohol, was that it stimulated the body while clearing the mind. Intellectuals now had discussions rather than orgies, and some coffeehouses, such as the Procope in Paris, served as centers of intellectual and artistic life where such men as Voltaire skewered aristocratic foibles.

The Café Heinrichhof in Vienna inspired Johannes Brahms and other great composers, as well as merchants who preferred the sound of money. Other coffeehouses (such as this author's grandmother's Café Mozart in Vienna) hosted cards and billiards and other such less-inspired diversions. The leisure of the coffeehouse was serious business.

Coffeehouses were also intimately involved in the birth of civil society and the democratization of semi-feudal aristocracies. Thus, it was in Paris's Café Foy that on July 13, 1789, Camille Desmoulins planned the assault on the Bastille that ushered in a new political age (Oliveira 1984). It is perhaps ironic that his action would also set into motion events that destroyed the coffee plantations of the world's greatest coffee producer, the French colony of St. Domingue (today Haiti).

Coffeehouses continued to be associated with subversion in the nineteenth century. Preparations for the 1848 revolutions were made in the coffeehouses of Berlin, Budapest, and Venice. In France, one of the first responses to threatened revolt was to close down the cafés (Barrows 1991). Ulla Heise (1987) argues, however, that although émigrés continued to

plot in cafés after 1850, the coffeehouses lost their revolutionary association. Revolution became more proletarian, and workers tended to frequent taverns, bars, and winehouses. S. Barrows, on the other hand, credits the declining political radicalism of the coffeehouses to the rise of newspapers, music halls, and other places for association.

If coffee in Europe was a part of the middle class's struggle for democracy, coffee itself became increasingly democratized. Although taxes kept the price high, coffee became the breakfast beverage of choice for the Continent's urban working class. To compensate for the high prices, the poor often drank coffee substitutes rather than the arabica (in Germany, *Kaffee* can refer to any number of beverages; the arabica or robusta is called *Bohnenkaffee*).

United States Consumption

As clanging factories gave birth to the industrial age, coffee came to represent labor as well as leisure. North American thirst was instrumental in making coffee a mass consumer product, a drug to prop up the drooping eyelids and awaken the flagging consciousness of an army of laborers. But this could occur only after they had abandoned tea. The citizens of the original 13 British colonies were tea drinkers, in part at least because British taxation and transport policies had made the arabica inaccessible to all but the rich. There were a few coffeehouses, such as Boston's Green Dragon, which Daniel Webster called the "headquarters of the revolution," and The Merchant's Coffeehouse in New York, also associated with the independence movement. But the taverns – which doubled as courtrooms and public meeting halls – were by far the favorite drinking spots in colonial times, although even those were little frequented because of the rural (and in places puritanical) nature of settlements.

United States coffee consumption in 1783 was only one-eighteenth of a pound per capita a year. By the 1830s, however, North Americans had cast aside tea for coffee, and by midcentury they were each drinking more than five pounds a year. Although the 1765 Stamp Act and the Boston Tea Party certainly dampened enthusiasm for tea, commerce and demography were probably more responsible for transforming the United States into the world's greatest coffee market. Resistance to the British East India Company was bolstered by the proximity of the French coffee-producing colonies of St. Domingue (Haiti) and Martinique and, later, Portuguese Brazil.

After independence, Americans were free to trade with these colonies, supplying them with slaves and naval stores in exchange for coffee. Thus, New England merchants introduced coffee into North America in relatively large volume. The price of coffee fell from 18 shillings per pound in 1683 to 9 shillings in 1774 to just 1 shilling in 1783. Such lower prices naturally

expanded demand, and government policy further aided the transformation as import taxes on the beans were first lowered and then abolished in 1832. Resumed during the Civil War, they were definitively abolished in 1872.

The absence of a coffeehouse culture in the United States meant that the beverage's popularity was not associated with political subversion as it was in the Middle East and Europe, and its consumption could be encouraged with no political danger. The flood of northern European immigrants from coffee-drinking countries probably also contributed to the shift from tea. Soon coffee drinking became entrenched as a social institution, and annual per capita consumption ballooned from under one pound at independence to nine pounds by 1882.

Nonetheless, it took time for the coffee trade in the United States to become institutionalized. It was originally specialized and artisanal; U.S. importers became involved only after the green beans reached New York. There, they sold them to wholesalers who, in turn, peddled them to thousands of retailers. As in Europe, there was no brand identity early on, with each grocer selling his own blends. But to a much greater extent than Europeans with their cafés, nineteenth-century North Americans bought beans in bulk at the grocers and roasted them at home.

This habit gave the housewife discretion over coffee purchases. Unlike her British sisters, who denounced coffee because of coffeehouses, the North American wife was the one who bought coffee; and brewing a good cup of coffee was often a measure of wifely abilities as a "homemaker." Consequently, in contrast to the customs in the Middle East or Europe, coffee merchandising in the United States became much more oriented to women than men. Arbuckle's 1872 advertisement, the first handbill in color for coffee, shows two women by the kitchen stove, the first complaining "Oh, I have burnt my coffee again," and the second counseling "Buy Arbuckle's Roasted, as I do, and you will have no trouble" (Ukers 1935: 451).

In fact, because so much coffee was purchased at the grocery store, where product differentiation was accentuated (rather than at the café where there was much less variety and brand was not displayed), coffee brand identification first arose in the United States. This was achieved through great efforts to create brand loyalty, although instead of market segmentation and differentiation, there was a tendency toward homogenization.

Today, to attract women to its brands, General Foods, one of the largest coffee merchandisers in the world, brings together hundreds of women in focus groups each year and sends out thousands of questionnaires. The coffee market is still female-oriented in the United States, as is demonstrated by the tendency to employ women in televised coffee advertisements.

Before brand loyalty, however, there came new technology to improve the quality and marketing of

coffee. In the nineteenth century, because of fairly primitive transportation and packaging techniques, the quality of coffee once it reached a Cincinnati or an Omaha was fairly poor. Consequently, consumers in such areas had a different idea of how coffee should taste than consumers do today. Some brewed it with an egg to give the drink a rich yellow color. It was also popular to add the uncooked skin of a mild codfish to the pot. In the western states, coffee was put in a saucepan and simmered for two to three hours and reheated as necessary. Mass-consumed coffee was not necessarily good coffee, but the cowboy huddled around the campfire was not a demanding gourmet. Under these conditions, it is not surprising that importers were slow to improve the quality of the coffee they sold. Roasted coffee could not be widely marketed because it lost its taste, and ground roasted beans lost flavor much more quickly. The green bean, on the other hand, could be stored for months or years without harm.

The demand for improved and standardized coffee probably derived in large part from improved transportation, roasting, grinding, and brewing technologies. Although green beans traveled well, they were frequently damaged during long sea voyages aboard sailing ships. Merchant efforts to dye damaged beans with rust, or indigo, or beef blood, or to glaze them with eggs improved their appearance, but not their flavor. Nor did the practice of roasting impair beans with cinnamon, cloves, cocoa, and onions help much. In addition, coffee grinders did not have sufficiently sharp blades to grind the beans finely, and using mortar and pestle was a laborious business. Consequently, grinds had to remain in contact with water longer to impart their flavor, which had the drawback of creating a bitter brew: Tannin, which causes the bitter taste, begins to be extracted from the grinds about 45 seconds after contact with hot water. And finally, coffeepots were primitive. Before the invention in 1800 of a kind of percolator with a built-in filter, most people simply threw grounds into water, much as with tea.

Many of these difficulties in brewing good coffee were resolved somewhat in the nineteenth century. Railroads sped coffee from the fields to ports where the rapid and relatively large steamships that replaced sailing ships spared the green beans ocean damage. Improved control was achieved over oven temperatures, allowing for more regular roasting by, for example, the spherical roaster invented by an Austrian, Max Bode, in 1851, and the pull-out roaster produced by a New Yorker, Jabez Burns, in 1864. Better grinders, producing finer and more even grounds, were also invented, and a welter of coffeepots were mass-produced. The first predictable pumping percolator, patented in France in 1827, became the most popular North American pot in the first half of the twentieth century. Drip pots were improved with the invention of the disposable filter in 1907 (the prototype, which

was cut from an ink blotter, was a considerable improvement over the previous horsehair filters). These pots were more popular in Europe, as was the espresso pot, first designed in 1837. Such pots were true monuments to the industrial age with its drawbacks and its charms. Espresso pots sometimes exploded. But if they did not, they could do a host of other things (Bersten 1993). One pressure pot also boiled eggs. The Armstong Perc-O-Toaster toasted bread and baked waffles while it perked coffee (Panati 1987).

Improvements in technology led to standardization, and eventually, to a wholesaling oligopoly in the United States. Whereas the French concentrated on devising new pots to improve the quality of brewing for the refined palate, North Americans focused on roasting, packaging, and marketing to reach the mass market. The first packaged roasted coffee was "Osborn's Celebrated Prepared Java Coffee," which appeared in 1860. The first brand to enjoy a national market was "Arbuckle's Ariosa," beginning in 1873. Sales of packaged coffees, however, were slow to replace those of green coffee beans until 1900, when Edwin Norton invented vacuum packing, which allowed roasted, ground coffee to retain its flavor. In 1903, Hills Brothers was the first company to commercially employ the process (Ukers 1935; Uribe 1954).

The ability to preserve roasted coffee in vacuum packages allowed a few national brands to dominate the trade in the United States. (Europeans were much slower to buy canned coffee, preferring beans and cafés). North American firms began to integrate vertically, with the A & P grocery chain in the forefront, even to the extent of stationing buying agents in the interior of Brazil, as well as importing, roasting, packing, and retailing its own brand. Still, many smaller brands and wholesale grocers persisted. In 1923 there were 1,500 roasters and 4,000 wholesale coffee grocers in the United States (Ukers 1930).

The market power of a relatively small number of powerful wholesalers combined with U.S. government policy to create standardization. Since its inception, the wholesale market had been completely unregulated and subject to rampant speculation and fraud. Many traders mixed together a number of coffee qualities, adulterated the mix with chicory and grains, and then called it "Mocca" or "Java." (The name "mocca" derived from Yemen's main port and stood for authentic coffee, not a combination of chocolate and coffee as it does today.) In the 1860s, a typical 122-pound bag of "Java" coffee arrived from Jamaica with about 5 pounds of sticks and stones added to it.

The freewheeling coffee market began to change in 1874 when a submarine cable tied South America to New York and London by telegraph. Information about prices and demand and supply became internationally homogeneous. The establishment, in 1882, of the New York Coffee Exchange, which was instituted to prevent commercial corners from driving up prices (as had happened in 1880), institutionalized

access to information, and Le Havre, Hamburg, and London followed with their own major coffee exchanges. Prices and grades thereby became more generalized, and coffee became more purely a commodity as well, in the sense that coffee shipments were now bought and sold on the market floor without the buyer actually seeing the lot in question.

Until the early twentieth century, professionals would judge a sample bean's quality on its color, size, shape, and shine. Later, taste tests were instituted to check aroma, body, bitterness, and richness. Coffees became commodities possessing a bundle of specific, graded attributes. Indeed, with the advent of futures, buyers purchased coffee not yet blossoming on distant trees. By 1880, merchants were already buying an idea, rather than palpable beans: In that year there were 61 million bags bought and sold on the Hamburg futures market when the entire world harvest was less than 7 million bags!

Grinding did not become standardized until 1948 when the National Bureau of Standards of the Department of Commerce issued guidelines for three categories of grind – regular, drip, and fine – which most U.S. producers still follow. Much earlier, however, the Pure Food and Drug Act had decreed in 1907 that imported coffee be marked according to its port of exit. Thus “Santos” became a specific type of coffee, as did “Java” or “Mocca.” There were more than a hundred different types of coffee imported into the United States, representing the greatest variety in the world. Importers were now less able to adulterate and defraud buyers.

The buyers, in turn, became more conscious of the quality of their coffee as they were able to buy professionally and uniformly roasted beans. The almost 4 million pamphlets issued by the National Coffee Roasters' Association at the beginning of the twentieth century in a campaign to educate housewives in proper brewing techniques apparently paid dividends. North American per capita consumption almost doubled between 1880 and 1920 to 16 pounds per capita. The growth of cities and factories accelerated the trend. No longer primarily the beverage of spiritual contemplation, commerce, or leisure, coffee became the alarm clock that marked industrial time. North American coffee imports swelled almost ninetyfold in the nineteenth century.

Temperance societies in the United States and Europe began to promote coffee and coffeehouses as the antidote to the alcoholism of the saloon, which was quite an ironic shift from the Islamic mullahs' fear of the brew's intoxicating effects. A sign in one Christian café read: “Coffee-house – God's house; Brandy shop – Devil's drop” (Heise 1987: 227). But there seems to have been no close relationship between coffee and alcohol. Coffee consumption did not suddenly increase in the United States with the onset of Prohibition, nor did consumption sharply drop with the relegalization of alcohol.

In another ironic twist, at the same time that the prohibitionists were singing coffee's praises, makers of cereal-based beverages launched an expensive attack on coffee's harmful properties. Coffee producers responded with an even more expensive defense of their drink. These mass-media campaigns encouraged the oligopolization of the market, and in 1933 just two companies, Standard Brands and General Foods, accounted for half of coffee's \$6 million outlay to advertise coffee on the radio.

Despite the “many bugaboos raised by the cereal sinners,” coffee became increasingly linked to sociability as ever more was drunk in public places (Ukers 1935: 477). The twentieth century saw the rise of the coffee shop and the cafeteria. Workers, especially white-collar workers needing to pause from their labor and socialize with their colleagues, took coffee breaks. Indeed, the beverage became embedded in popular speech. “Let's have a cup of coffee” came to mean “let's have a conversation.”

Restaurants began using coffee to attract customers by keeping the price low and offering unlimited refills. (Iced tea was the only other beverage to be given this privileged status until the recent inclusion of soft drinks). Grocery stores often employed coffee as a loss leader to bring in shoppers, and when in the middle 1970s prices rose steeply, many grocers absorbed the higher price rather than alienate their customers. Because of its strong connection with sociability, its tendency to addict consumers, the medicinal effect of its caffeine, and the small number of substitutes, coffee has come to be viewed as a necessity more than almost any other food or beverage (Lucier 1988; Oldenburg 1989).

The growth of the vast U.S. market for coffee and the beverage's privileged social function led to the expansion both vertically and horizontally of a few companies creating an oligopoly. Today three companies – General Foods, Proctor and Gamble, and Nestlé (which also dominates much of the international market) – are responsible for 80 percent of the U.S. coffee market. They spend hundreds of millions of dollars a year to promote their brands, yet paradoxically, the price and profit levels of coffee remain low, despite the drink's status as a necessity, even a drug, for which the taste is perhaps less important than the effect it produces and the price it commands.

In recent years, however, specialty coffee beans, sold mostly by small companies and cafés, have challenged the conventional wisdom that North American consumers are unwilling to pay a high price for good coffee. Gourmet coffees, offered by creators who stress their national origins, the roast employed, and sometimes the flavorings added, have collectively become the only sector of the U.S. coffee market that is growing. They tend to appeal to younger, more affluent buyers for whom gourmet coffee is more a status symbol or a declaration of one's lifestyle than it is a necessity.

Yet even before the rise in the popularity of gourmet coffee, roasters had sought ways of expanding the mass market. After many attempts, the first commercially successful dried coffee was produced in 1906 by a North American chemist residing in Guatemala. But it attracted few drinkers until World War II when it was included in soldiers' rations, and since that time the market for instant coffee has grown substantially. By the 1960s, as much as one-third of home-prepared coffee was instant soluble. Its ease of preparation helped expand consumption but undermined quality because instant coffee utilizes mostly robusta coffee, which is a faster-growing but more bitter species than the arabica. As with gourmet beans, since the 1980s there has been a trend toward adding other flavorings to soluble coffee to produce specialty drinks that resemble, to name a few, Irish coffee or cappuccino.

Another major innovation has been decaffeinated coffee. It was developed in Germany at the beginning of the twentieth century by Ludwig Roselius, a sworn enemy of caffeine, which he blamed for his father's death. In Roselius's original process, green coffee was steamed and then soaked in a chlorinated organic solvent. Other processes have subsequently been developed, some involving the breeding of coffee trees with low caffeine yields. As coffee has come under attack for contributing to cardiac problems, decaffeinated consumption has soared.

The medical community is divided on the effects of coffee drinking. There are certainly beneficial effects, and the brew is sometimes prescribed in the treatment of barbiturate poisoning, migraines, chronic asthma, and autism in children. That heart ailments may be a negative effect is strongly debated, with each side marshaling impressive evidence. It seems clear that coffee's physical effects vary greatly, depending on the consumer. One study has concluded that for about 14 percent of the population, caffeine dependence produces a physical addiction similar to an addiction to alcohol or cocaine.

Such controversy over the ill effects of coffee helped drive down per capita consumption in the United States from its peak of 3.2 cups per day in the 1960s to 1.8 cups in 1993. Still, health concerns alone cannot explain this retreat because consumption of two other beverages with caffeine – soft drinks and tea – has grown since 1970, with soft drinks more than doubling in per capita consumption to reach almost 40 percent of all beverages consumed. By contrast, milk consumption has declined and that of juices has remained flat (United States Department of Agriculture 1993). No doubt advertising and packaging have had a substantial impact, and it is the case that soft drinks and bottled water (another rapidly growing beverage) require no preparation.

Numerous coffee companies rose to the challenge to compete directly with soft drinks by creating new products, such as prebrewed coffee, bot-

tled iced coffee, and iced cappuccino. This seemed to be the logical terminus of a century-long process in which the activities of roasting, grinding, and brewing, formerly done in the home, became industrialized.

Yet there is a countertrend as well in the growing gourmet market. The swelling army of connoisseurs who want to make American coffee "a national honor," to borrow the words of coffee expert W. Ukers written decades ago (1935; 570), rather than a "national disgrace" are buying a great variety of specialty roasted beans and grinding them at home. Specialty coffeepots and espresso makers also constitute a booming market.

Coffee Drinking in the World

In other parts of the world, the popularity of coffee is at least being sustained or is on the increase. Coffee-producing countries have increased their share of consumption from less than 10 percent of production at the end of the nineteenth century to about one-quarter by the end of the twentieth century. Among coffee growers, Costa Ricans have the greatest taste for their own beans, and Brazilians are second. But even in these nations, per capita consumption is well under one-half that of northern European countries. Africans, except for Ethiopians, consume almost none of the beans they produce. Coffee is still very much a commodity consumed in rich countries and produced by poor ones, and per capita consumption is closely correlated to the wealth of consuming nations. It is also negatively correlated to an ability to grow arabica bushes or to historic ties to the trade (United States Department of Agriculture 1993). Few of the countries earliest involved in coffee's history – Yemen, Turkey, Indonesia, Haiti, or Martinique – are significant consumers, and in Europe, those with the oldest ties to the coffee trade – Greece, Portugal, Italy, and the United Kingdom – are among the lowest consumers. On the other hand, countries in Asia with no historic connection with the arabica – Japan, Korea, and Singapore – are rapidly increasing per capita coffee drinking as incomes climb. In 1992, Japan was the fourth largest coffee importer in the world. In mainland China, the arabica is making inroads not as a proletarian beverage but as a bourgeois one.

Worldwide, the human race drank about 380 billion cups of coffee in 1991¹ or about 76 cups for each man, woman, and child on the planet. Clearly, the exotic drink that Yemenis first tasted nearly 600 years ago has assumed a position of some considerable global importance.

Production in Arabia

Despite the fact that *Coffea arabica* grew wild in Ethiopia and *Coffea robusta* appeared naturally in the Congo, Yemen had a virtual world monopoly on pro-

duction for about half of coffee's 600 years of lifetime as a commodity. Stern measures were taken to prevent the smuggling out of coffee plants, and although this move was not entirely successful – there are reports of coffee growing on India's Malabar coast and perhaps Ceylon in the sixteenth century – Yemen's grip remained firm until the middle of the eighteenth century. In the early part of that century, Yemen may have been producing some 20 million pounds a year (Raymond 1973–4; Becker et al. 1979).

The beans were grown in small, irrigated mountain gardens in various small, broken areas of northern Yemen, then transported by camel either to port or across the deserts (Daum 1988). The entire crop was consumed in the Middle East and Southwest Asia until the middle of the seventeenth century, and even thereafter, the East remained the main market for another hundred years.

Europeans initially purchased Yemen's coffee through Arabian traders. The Dutch and British East India Companies eventually established factors in Mocca, but they still exercised no control over production and had to make purchases with gold and silver because the Arabs much preferred Asian trade goods to European products. Another drawback was that such wealth attracted pirates, often based in Madagascar. Not satisfied with this precarious trade in a product whose value was growing vertiginously, the Europeans smuggled out coffee plants to begin production elsewhere. This decision ultimately spread the Arabian bean to more than 100 countries on 4 continents. Yemen soon became an inconsequential coffee producer, and the thriving coffee port of Mocca, which probably had more than 30,000 inhabitants at its height, dwindled to 400 stragglers living amidst its ruins in 1901.

Production in Asia

More than any other commodity, coffee was produced in poor colonies and countries for the enjoyment of those in rich countries. But, as we have seen, the Europeans who created coffee colonies (the Dutch, British, French, and Portuguese) were not the ones who consumed the most coffee on a per capita basis. Coffee was not only an export crop but a reexport crop.

In 1690, the Dutch introduced the coffee bush to Java, where their colonial might was soon employed to force peasants to labor on coffee plantations. Others were compelled to grow trees on village lands and give over shares or to sell at fixed low prices to government agents. Although clearly a coercive system, it relied on traditional local power relations and peasant agriculture to extract profit. Technology was primitive and yields low. But costs, for the Dutch, were even lower; so profits were high. The Javanese, however, experienced little capital accumulation or economic development (Kok 1864; Geertz 1971). A different

kind of labor system was employed on the tiny island of Bourbon (renamed Reunion), lying southwest of Madagascar, where French colonists forced African slaves to grow coffee. In the eighteenth century, the island became one of the world's largest coffee producers.

Throughout Asia, coffee growing was a colonial enterprise as it continues to be in Papua New Guinea (Stewart 1992). Dutch and British East Indian colonies were among the world's leading producers until the last part of the nineteenth century when a fungus devastated coffee fields. And disease has plagued coffee growers the world over because "coffee is one of the tropical plants most susceptible to diseases and insect attacks which may destroy whole plantations" (International Institute of Agriculture 1947: 22).

Many East Indian plantations turned to tea or rubber, and where coffee maintained its hold, as in Sumatra, it was mostly the robusta, rather than the disease-plagued arabica, that was grown on peasant plots. Ceylon (Sri Lanka), India, and the Philippines were also major coffee producers into the nineteenth century, but were overwhelmed by Latin American production in the twentieth century.

Production in the Caribbean

The Dutch, who brought some arabica trees to Amsterdam's botanical garden, assisted coffee in its move westward by planting it in Surinam; later some seedlings were transported to Brazil. Another early Dutch contribution to the dissemination of coffee in the Americas was even more circuitous. An arabica tree, given by the mayor of Amsterdam to Louis XIV, was cultivated in Paris's botanical garden. One of its seedlings, however, made its way to the Americas early in the eighteenth century when the Frenchman Gabriel de Clieu carried it across the Atlantic to a New World home in Martinique. From there it spread to St. Domingue (today Haiti) and later the mainland.

On St. Domingue, colonial production was perfected when slaves, already imported to grow sugar on the island, were also employed in the coffee fields. European consumption grew tenfold in the 50 years between 1739 and 1789, with French colonies supplying three-quarters of the total. In fact, relatively inexpensive Caribbean coffee was already displacing Yemeni beans even in the Cairo market, demonstrating, among other things, the competitive advantage of slave labor and plantation agriculture over the garden plots of Yemen and the peasant farms of Java. Nonetheless, the arabica remained a rather exceptional specialty product under mercantilism. In 1800, on the eve of the Industrial Revolution, Europeans consumed on average only about three-quarters of a pound per year.

By the end of that century, however, coffee had become the world's third most important traded commodity in terms of value, thanks to the Industrial Rev-

olution and its transformation of transportation systems and markets. Perhaps fittingly, it was also during the nineteenth century that coffee was grown for the first time on an industrial scale. This occurred in the fields of Latin America where, unfortunately, coffee helped to sustain slavery – even to resurrect it.

The slave revolution in St. Domingue severed the colonial tie with France, abolished slavery, gave birth to modern Haiti, and sent the island's coffee economy into an irreversible decline. But rather than ending slavery, French and Spanish planters made Cuba the next great slave-operated coffee system. Its reign, however, was short-lived because the profitability of sugar overshadowed that of coffee. Meanwhile Brazilians, who were having difficulty competing with Cuban sugar, switched to coffee growing, and Brazil in the 1830s began a domination of world coffee production that has endured to this day. For most of the nineteenth century, Brazil owed that domination to the toil of slaves.

Brazil

The transition from slave-based sugar to coffee growing in Brazil was natural, but not inevitable. An agricultural downturn during the first four decades of the nineteenth century propelled government officials, and even some planters, to reassess the newly independent country's export orientation and reliance on human chattel. But a burst of European and North American demand for coffee redoubled reliance on the export economy and stimulated the craving for slaves; the planter elite, once again, confidently pronounced Brazil an "essentially agricultural country" and, more specifically stated, that Brazil *was* coffee, and coffee meant slaves. The rate at which Africans were landed in Rio de Janeiro and Bahia between 1800 and the abolition of the Atlantic trade in 1850 far surpassed that of any previous place or time. Brazil had been by far the largest importer of African bondsmen and was the last country of the Western world to abolish slavery when emancipation finally arrived in 1888 (Curtin 1969).

Contemporary critics complained that slavery was not only socially dangerous but was also delaying Brazil's social, economic, and political development. Later, in the 1960s, a school of analysis known as "dependency theory" extended this critique. The coffee industry was accused of reinforcing slavery, which, in turn, had impeded the development of a domestic bourgeoisie, restricted commodity and capital markets, skewed wealth distribution, created an unfavorable view of manual labor, and fashioned a liberal oligarchic state that sold national sovereignty and progress in the bags of coffee shipped abroad (Frank 1969; Santos 1974). Indeed dependency theory maintained that not only slavery but also monocultural export economies in general had led to the underdevelopment of neocolonies that were tied to the world economy.

A counter-argument, however, concedes that the *dependistas* were correct in their assessment insofar as the Paraíba Valley (parts of Minas Gerais, Rio de Janeiro state, and São Paulo state) was concerned, but claims that they were wrong in the case of western São Paulo where coffee stimulated capitalist development (Amaral Lapa 1983). There, when emancipation became inevitable, planters turned to immigrant labor. Some 3 million southern Europeans emigrated to Brazil between 1880 and 1930, most of them headed toward São Paulo's coffee fields.

These families of workers helped Paulista planters become agro-industrialists on a scale previously unknown in coffee cultivation. Indeed, Paulistas established some of the largest plantations (*Fazendas*) ever built anywhere, at any time. The Cambuhy Estate, for example, spread out over 250,000 acres on which almost 3 million trees were grown, with the whole tied together by 60 kilometers of private railroad track and 300 kilometers of roads.

These Paulista agro-industrialists directed some of their agricultural capital not only to urban real estate, public works, and government bonds but also to railroads, banks, and even factories. The *fazendeiro* was thus transformed from rentier to capitalist, from coffee baron to entrepreneur. He became the leading partner in what is generally acknowledged to be the most progressive national bourgeoisie in Latin America – indeed, one of the most entrepreneurial in the entire developing world. He industrialized São Paulo, and coffee was transformed from a colonial product to the foundation of the nation-state (Dean 1969; Cano 1977; Silva 1981; Levy 1987).

Albert Hirschman (1981) has argued that coffee had special advantages that encouraged the development of a national bourgeoisie and industrialization. He points out that the relative lack of forward and backward linkages, and the simple technology sufficient to grow and refine coffee without great capital requirements, may have stimulated entrepreneurial initiative by keeping foreigners restricted in their participation to the areas in which their comparative advantages lay, namely commerce and transportation. Thus, unlike petroleum, copper, or even sugar, the production and processing of coffee (except roasting and grinding) was done almost exclusively by nationals. Agricultural production profits remained within Latin America and because coffee did not demand great infusions of capital, planters were free to diversify into other areas.

The economic wealth of planters translated into political power which, when wielded to protect their own interests, led to a fundamental transformation of the liberal state and, ultimately, of the world market for coffee. Beginning with the valorization of coffee in 1906 through the creation of the Institute for the Permanent Defense of Coffee in the 1920s and, finally, the *Departamento Nacional de Café* in 1933, the

Brazilian federal and state governments came to finance most of the world's coffee trade and hold most of its visible stocks. Seven years later, one of the first international commodity cartels was established with the Inter-American Coffee Agreement. And in 1962, the rest of the world's producers were brought under the umbrella in the International Coffee Agreement, which persisted until 1989.

Coffee set the precedent that other raw-material producers would later follow, including those of OPEC. It also transformed the state's role in the domestic economy. By the end of the First Republic in 1930, the Brazilian state was responsible for much of the financing, warehousing, transportation, and sale of coffee and controlled one of the world's largest commodity markets. Coffee, thus, carried Brazil from an archaic slavocratic social formation to state capitalism in half a century (Topik 1987), and the state played a similarly large economic role in most other coffee-producing countries.

Spanish-American Production

In other coffee-producing lands, large-scale plantations like those of Brazil were unusual even during the height of the export boom. In Venezuela, for example, despite a tradition of slave labor and haciendas, smaller-scale production eventually prevailed (Roseberry 1983). In Colombia, the eastern part of the country had a strong colonial tradition of peasants, with no slaves or latifundia, whereas in the west there had been both. The east was the leading growing area until the beginning of the twentieth century when western provinces, such as Cundinamarca and Antioquia, rapidly increased production. But in Cundinamarca, coffee was grown on shares; only in Antioquia were haciendas important social units, and there were also many small growers as well. Nationally, by 1932, 87 percent of all coffee farms had less than 5,000 trees (farmers usually placed 500 to 800 trees per acre), and 74 percent of all production came from farms of less than 20,000 trees, meaning less than 40 acres (Beyer 1947; Nieto Arteta 1971; Palacios 1980).

In Central America, coffee was also grown mostly on small and middle-size farms called *fincas*. Although Costa Rica was far from being a rural democracy (71 percent of the peasantry was landless in 1883) and land ownership was concentrated, even the biggest coffee plantations were often discontinuous and fragmented into a number of small or medium-size lots, sometimes several kilometers apart (Cardoso 1977: 175–6; Seligson 1980). In fact, even the “great estates” in Costa Rica were defined as anything over 76 acres, which in Brazil would scarcely have been considered a middle-sized holding.

Elsewhere in Central America, coffee estates were somewhat larger than in Costa Rica. Nicaragua did have some coffee haciendas, but they were neither numerous nor extensive. El Salvador, the most depen-

dent of all countries on coffee, had a dynamic agrarian bourgeoisie with extremely efficient production techniques and concentrated land holdings. But production was intensive, not extensive as in Brazil. Eugenio Aguilar was considered a large, wealthy planter with two *fincas* that held 230,000 trees on less than 300 acres. But in Guatemala and southern Mexico, with larger frontier areas, there were substantially larger holdings. Many were spread over 5,000 acres in Guatemala, and in Tehuantepec, Mexico, there were several coffee estates with 1.5 million trees on some 20,000 acres (Lindo-Fuentes 1990).

Coffee estates, in general, were smaller in Colombia and Central America than in Brazil because of terrain and transportation problems, along with a shortage of capital and labor, the relative absence of a frontier, and the presence of people with preexisting land claims. To compensate for the higher production costs wrought by such problems, growers had to produce a better coffee that would command a higher price at market.

Higher-quality coffee required more labor-intensive harvesting. Shade trees had to be planted, and the berries were picked individually just as they ripened, in contrast to the procedure of stripping entire branches as was done in Brazil. Workers would sometimes have to pass the same tree six or seven times during the harvest instead of just once. Better quality also demanded the more sophisticated and expensive “wet” method of treating the berry.

While inspecting this method in Costa Rica, the Brazilian planter Jorge Dumont Villares marveled that the preparation of coffee was a “true science” there. The berries were picked, hulled, then left 24 to 40 hours to ferment. Next, they were washed to remove the outer membrane and dried for one or two days. From the drying grounds, the beans were taken to dryers where they were left for 25 to 35 hours to dry. The parchment film was then removed mechanically and the beans sorted by size, quality, and color. Finally, the beans were polished (Villares 1927). Producing high-quality coffee, then, demanded substantial capital and technical expertise, as well as abundant labor. The capital was provided more by mill owners than by growers.

The relationship between coffee and the kind of labor applied to it varied from country to country and changed over time. As noted, coffee reinforced slavery in the Caribbean and in Brazil, but slavery was not common elsewhere in Latin America where other coercive forms of labor were applied. Nonetheless, as we have tried to show, coffee, plantations, and slavery or forced labor were not necessarily linked, and coffee did not always yield a strong landlord class. Rather, in Costa Rica, for example, merchants and millers came to predominate because of the small-holding pattern, and elsewhere merchants also became the dominant class. In Colombia, most of the progressive planters were initially merchants, and

only in such areas as Cundinamarca were traditional landlords the coffee growers. Many Nicaraguan *finqueros* were urban based, and the El Salvador coffee-grower class was largely composed of merchants and other urban-based capitalists who invested in land in the 1880s and 1890s as coffee prices rose. German merchants in Guatemala and Chiapas followed the same path.

In few cases was there much continuity between the colonial landed elite and the coffee bourgeoisie. In fact, many members of the coffee elite were nineteenth-century immigrants. Thus, it is a mistake to refer to the coffee growers of Central America and Colombia as *Junkers* or to discuss "landlord capitalism" as is often done (Winson 1989). Coffee was developed by a new bourgeois group that was not reluctant to invest in other sectors of the economy as well.

This does not mean that there was no resistance to the rule of coffee. In Guatemala, villagers frequently destroyed coffee *fincas* (Cambranes 1985). Peasants in western El Salvador, the center of coffee cultivation, in collaboration with the Communist party, attempted to overthrow the government in 1932. The result was at least 17,000 peasants dead in the gruesome massacre known simply as *La Matanza* (Anderson 1971; North 1985).

The willingness of coffee growers to use violence has often allowed the essence of their rule to continue while changing its appearance. Colombia never had a truly populist ruler; the traditional Liberals and Conservatives still share power. El Salvador and Guatemala have the most ferocious and brutal regimes of Central America, still overseen by descendants of the coffee elite. São Paulo was the last state to hold out against the populist leader Getúlio Vargas, even waging civil war against him in 1932.

Yet there has been change. São Paulo, for example, although its economy is now based on the cities and factories that coffee gave birth to rather than on the countryside, has become the home of Latin America's largest socialist party, the *Partido dos Trabalhadores*. In Costa Rica, a revolution in 1949 brought to power a semiautonomous state that has presided over Latin America's most vigorous democracy; and in Nicaragua, division among growers and a single family's attempt to monopolize wealth led to the creation of the first truly socialist regime on the American mainland. In all of these cases, however, democratic trends began to grow only *after* coffee ceased to be the dominant export.

Socioeconomic Aspects and a Return to Africa

During the years 1850 to 1930, the world coffee market was fairly homogeneous, with the vast majority of the world's coffee (as much as 95 percent) produced in Latin America. Although there are an estimated 50 different species of the genus *Coffea*, and a much

greater number of small species or elementary hybrid-producing forms, over 90 percent of exports before World War II were of the same species, the arabica.

Such apparent uniformity, however, is a bit misleading. Because all coffee is cultivated outside of its natural habitat, under different methods and a wide range of natural conditions, the arabica is different on virtually each plantation (International Institute of Agriculture 1947: 19–20). Beans also differ from year to year. There are 10 generally recognized subspecies.

The international market divided coffee by port of exit and grade of the bean (determined by color, size, and degree of breakage). The main division in the world economy, however, was between "Brazils" and "milds"; the latter, produced in Colombia and Central America, began making inroads into the dominance of the former beginning in the 1920s, when Brazil's coffee valorization program propped up the price of the stronger and more bitter "Brazils." Different national crops also varied substantially according to the harvest season. But the time it took for beans to reach market was not of great consequence because they did not deteriorate over time (as, for example, apples, did) and importers maintained large stocks in the consuming countries to smooth the flow to market.

Despite such differences, the producing countries shared many similarities. They exported almost all of their crop to Europe and North America. And all used fairly similar cultivation techniques demanded by the arabica bush or "tree." In order to increase yield and facilitate the harvest (done almost always by hand), the trees were topped in their third year and only allowed to grow to 6 to 16 feet, though in the wild they grow to 35 feet; hence they were often termed bushes rather than trees. The arabica grows best with deep, permeable, well-aerated soil and a minimum rainfall of 50 to 60 inches during germination (coffee lands in Latin America were almost never irrigated as they were in Yemen and Java). It also needs a long, dry season and mild climate, with the optimum temperature between 59 and 77 degrees Fahrenheit.

Coffee-producing countries have also shared a secondary position in the industrialization of coffee. Since the consumers were in the rich countries, the retailing, wholesaling, roasting, grinding, and packaging were done in those countries. Growers were, thus, excluded from the most technologically sophisticated aspects of the production of coffee, from which the greatest innovations such as packaging, vacuum sealing, and soluble coffee emerged. These, along with roasting and advertising, have also constituted the most lucrative of the forward linkages. Thus the Santos price was at most half the New York retail price, and if the coffee were sold by the cup rather than the can, the additional labor cost and markup meant that as much as 90 percent of the value added was created in the consuming countries.²

Today the supermarket price of standard roasted or ground coffee is 3 to 4 times the New York green

price. Restaurants sell coffee by the cup for 6 to 10 times the retail bean price. And the specialty brands are even more expensive. Thus, upwards of three-quarters of the value is added in the consuming countries. The only dent in the consuming countries' market command has been made by the increasing ability of some of the major exporting countries to produce and export soluble coffee.

For most of the coffee-producing Latin American countries, the cultivation of coffee became important toward the end of the nineteenth century or early in the twentieth century, and coffee came to dominate national exports, creating monocultures. In 1929, coffee cultivation was responsible for 71 percent of exports in Brazil, 61 percent in Colombia, 77 percent in Guatemala, 77 percent in Haiti, 54 percent in Nicaragua, 67 percent in Costa Rica, and fully 93 percent in El Salvador (Ukers 1935).

Since World War II, however, coffee cultivation has slowly shifted back to Africa. Although Latin America still produces most of the world's coffee, the crop is no longer responsible for more than half of the exports of any one of its countries. In Africa, however, in the late 1960s, coffee represented more than half the exports of Angola, Burundi, Ethiopia, and Rwanda, and almost half of Uganda's (Barbiroli 1970), and the continent now supplies about 18 percent of total world exports. Africans usually plant the more disease-resistant and faster-maturing robusta bush. It produces a less desirable brew but has been particularly successful in the soluble coffee market.

Coffee has come full circle in Africa. Ethiopians had vast forests of wild arabica trees but did not enter the trade until this century. Indeed, a nineteenth-century Dutch observer wrote that "the plant is not only no object of culture there, but is abandoned to persecution in some parts through the rage of blind fanaticism" (Kok 1864: 210).

European colonial regimes and planters initiated the coffee export economy in Ethiopia, as well as in Kenya, Angola, Uganda, and elsewhere. Africans were sometimes pointedly excluded by law from growing the arabica or robusta, although in Zaire and Ethiopia they were forced to grow or harvest the beans. In the last 40 years, however, decolonization has transformed the situation. Today, the Ivory Coast, Ethiopia, and Uganda are the principal coffee-growing nations of Africa, and African peasants, not European plantation managers, are the overwhelming producers (Clarence-Smith 1994).

Conclusion

Throughout its history, coffee has stimulated ideas, debates, commerce, and development, not to mention numbing people to routine and exposing them to vice and exploitation. And it has helped to subvert cultures, social systems, and governments. In the consuming countries, coffee moved from the mystical

and mercantile to become one of the most traded bourgeois products in the world. Coffeehouses operated as centers of a bourgeois lifestyle for literati and businessmen alike, as well as meeting places for those who agitated for democratic politics. Coffee became the fuel of the industrial age.

In the fields of Latin America, however, European and North American demand led first to an intensification of slavery and then, in various places, to the appropriation of village lands, the expulsion of native peoples, and coerced labor. But, although there were large planters, there were also plenty of smallholders.

The story of coffee is clearly one of diversity. Geography, history, and local resistance combined to create a wide variety of social arrangements. To trace the history of coffee is to trace the path of the world economy over the last six centuries. From an Asian monopoly, to a European colonial product, to a global commodity grown on four continents, coffee has linked the different worlds of the producer and the consumer, the underdeveloped and the developed, the free and the enslaved, the rich and the poor, and the bourgeois and the archaic. Sufi Sheikh, sitting in the shade in Mocca, had no idea what a global force he was putting into motion when he sipped the first cup of coffee.

Steven C. Topik

The author would like to thank Gervase Clarence-Smith for his insightful comments.

Notes

1. According to the United States Department of Agriculture (1993), total world consumption in 1991 was 72 million sacks of 60 kilograms each. I multiplied by the recommended 40 cups of coffee to the pound to arrive at the number of cups.
2. This estimate assumes the New York raw material to be 50 percent of the New York wholesale price. It also assumes a 50 percent markup to the Midwest, and a 25 percent profit for the wholesaler and for the retailer. Finally, it reflects the conservative estimate that a retail cup of coffee costs the consumer about twice one made at home. A restaurant cup of coffee costs 7 to 13 times the retail price of coffee.

Bibliography

- Amaral Lapa, J. R. do. 1983. *A economia cafeteira*. São Paulo.
- Anderson, Thomas P. 1971. *Matanza: El Salvador's communist revolt of 1932*. Lincoln, Nebr.
- Barbiroli, G. 1970. *Produzione e commercio internazionale del caffè*. Bologna.
- Barrows, S. 1991. 'Parliaments of the people': The political culture of cafes in the early Third Republic. In *Drinking behavior and belief in modern history*, ed. S. Barrows and R. Room, 87-97. Berkeley, Calif.

- Becker, H., Volker Hoehfeld, and Horst Kopp. 1979. *Kaffee aus Arabien: Der Bedeutungswandel eines Welt-Wirtschaftsgutes und seine siedlungsgeographische Konsequenz an der Trockengrenze der Okumere*. Wiesbaden.
- Bersten, I. 1993. *Coffee floats, tea sinks: Through history and technology to a complete understanding*. Sydney.
- Beyer, R. 1947. *The Colombian coffee industry: Origins and major trends, 1740-1940*. Minneapolis, Minn.
- Bramah, E. 1972. *Tea and coffee. A modern view of three hundred years of tradition*. London.
- Browning, D. C. 1971. *El Salvador; landscape and society*. Oxford.
- Cambranes, J. C. 1985. *Coffee and peasants in Guatemala: The origins of the modern plantation economy in Guatemala, 1853-1897*. Stockholm.
- Cano, Wilson. 1977. *Raízes da concentração industrial em São Paulo*. São Paulo.
- Cardoso, C. 1977. The formation of the coffee estate in nineteenth-century Costa Rica. In *Land and labour in Latin America*, ed. K. Duncan and I. Rutledge, 165-202. Cambridge.
- Carlier, O. 1990. Le café maure. Sociabilité masculine et effervescence citoyenne (Algérie xvii-xx siècle). *Annales E.S.C.* 45: 975-1003.
- Clarence-Smith, Gervase. 1994. "Africa and world coffee markets to 1914." Unpublished manuscript.
- Curtin, P. 1969. *The Atlantic slave trade: A census*. Madison, Wisc.
- Daum, W. ed. 1988. *Yemen: 3000 years of art and civilization in Arabia Felix*. Innsbruck.
- Dean, W. 1969. *The industrialization of São Paulo*. Austin, Tex.
- Ellis, A. 1956. *The penny universities*. London.
- Frank, A. G. 1969. *Capitalism and underdevelopment in Latin America*. New York.
- Geertz, C. 1971. *Agricultural involution*. Berkeley, Calif.
- Hattox, R. 1985. *Coffee and coffeehouses: The origins of a social beverage in the medieval Near East*. Seattle, Wash.
- Heise, Ulla. 1987. *Coffee and coffee-houses*, trans. Paul Roper. West Chester, Pa.
- Hirschman, Albert. 1981. A generalized linkage approach to development, with special reference to staples. *Economic Development and Cultural Change* 25, supplement: 67-98.
- Holloway, Thomas. 1980. *Immigrants on the land*. Chapel Hill, N.C.
- International Institute of Agriculture, Bureau of F.A.O. in Rome. 1947. *The world's coffee*. Rome.
- Kok, A. S., ed. 1864. *Colonial essays*. London.
- Leclant, Jean. 1979. "Coffee and cafés in Paris, 1644-1693," trans. Patricia M. Ranum. In *Food and drink in history*, ed. Robert Forster and Orest Ranum, 86-97. Baltimore, Md.
- Levy, D. 1987. *The Prados of São Paulo, Brazil: An elite family and social change, 1840-1930*. Athens, Ga.
- Lindo-Fuentes, H. 1990. *Weak foundations: The economy of El Salvador in the nineteenth century, 1821-1898*. Berkeley, Calif.
- Lucier, R. L. 1988. *The international political economy of coffee; from Juan Valdez to Yank's Diner*. New York.
- McCreery, M. G., and M. Bynum. 1930. *The coffee industry in Brazil*. Washington, D.C.
- Nieto Arteta, L. E. 1971. *El café en la sociedad Colombiana*. Bogotá.
- North, L. 1985. *Bitter grounds: Roots of revolt in El Salvador*. Westport, Conn.
- Oldenburg, R. 1989. *The great good place*. New York.
- Oliveira, J. T. 1984. *História do café no Brasil e no mundo*. Rio de Janeiro.
- Palacios, M. 1980. *Coffee in Colombia, 1850-1970; an economic, social and political history*. Cambridge.
- Panati, C. 1987. *Extraordinary origins of everyday things*. New York.
- Raymond, A. 1973-4. *Artisans et commerçants . . .* Damascus.
- Roseberry, W. 1983. *Coffee and capitalism in the Venezuelan Andes*. Austin, Tex.
1991. "La falta de brazos: Land and labor in the coffee economies of nineteenth-century Latin America." *Theory and Society* 20: 351-82.
- Santos, T. dos. 1974. "Brazil, the origins of a crisis." In *Latin America, the struggle with dependency and beyond*, ed. R. Chilcote and J. Edelstein, 409-90. New York.
- Seligson, M. 1980. *Peasants of Costa Rica and the development of agrarian capitalism*. Madison, Wis.
- Silva, Sergio. 1981. *Expansão cafeeira e origens da indústria no Brasil*. São Paulo.
- Stewart, R. G. 1992. *Coffee: The political economy of an export industry in Papua New Guinea*. Boulder, Colo.
- Topik, S. 1987. *The political economy of the Brazilian state, 1889-1930*. Austin, Tex.
- Ukers, W. 1930. *Coffee merchandising*. New York.
1948. *The Romance of coffee: An outline history of coffee and coffee-drinking through a thousand years*. New York.
1935. *All about coffee*. New York.
- United States Department of Agriculture, Foreign Agricultural Service. 1993. *World coffee situation*. Washington D.C.
- Uribe, C. A. 1954. *Brown gold: The amazing story of coffee*. New York.
- Villares, J. D. 1927. *O café e sua produção e exportação*. São Paulo.
- Wenner, M. W. 1991. *The Yemen Arab Republic: Development and change in an ancient land*. Boulder, Colo.
- Wickizer, V. D. 1959. *Coffee, tea and cocoa, an economic and political analysis*. Stanford, Calif.
- Winson, A. 1989. *Coffee and democracy in modern Costa Rica*. London.

III.5 ☞ Distilled Beverages

Alcoholic beverages have been a part of human culture since at least the Neolithic period. Yet until recently, beverages made from fruits, grains, or honey were considered to be what historian Wolfgang Schivelbusch (1992) has called "organic," meaning that the amount of sugar in the ingredients produced the amount of alcohol in the drinks. Examples of such beverages are beer and wine. Beginning in the period from about A.D. 800 to 1300, however, people in China and the West learned to distill alcoholic liquids. This chapter traces the history of distilled alcohol and discusses the nature of several kinds of liquor.

Distillation and Alcoholic Beverages

Distillation is a method for increasing the alcohol content (and, thus, the potency) of a liquid already containing alcohol – the existing alcohol content usually the result of the fermentation of vegetable sugars. The distillation process separates the alcohol from other parts of the solution by the heating of the liquid to 173° Fahrenheit, a temperature sufficient to boil alcohol but not water. The resulting steam (vaporized alcohol) is collected and condensed, returning it to liquid form – but a liquid with a much higher proportion of alcohol than before. Repeating the process increases the liquor's potency yet further. Because distilled alcohol contains bad-tasting and dangerous chemicals called fusel oils (actually forms of alcohol) and congeners, both by-products of the distilling process, it is often aged in a procedure, originating in the eighteenth century, that rids the beverage of these chemicals. As the liquid ages, its container (preferably made of wood) colors and flavors it to produce a smoother and better-tasting product (Ray 1974).

A constant theme in discussions of distilled liquor is that of fire, which has three different metaphoric meanings. First, beverages are “burnt,” or distilled, over the flame of a still. Second, although it is a drinkable liquid, distilled alcohol is capable of combustion. The third meaning is an apt description of the sensation experienced by consumers of distilled spirits. “Firewater,” *aguardente*, *aguardiente* (meaning rough or burning water), and ardent (burning) spirits are all terms referring to such a sensation (Needham 1984).

Stills are the traditional equipment needed to distill alcohol. There are many different types. The earliest known is the *ambix* (plural *ambices*) used by Greek alchemists. *Ambices* were ceramic or metal pots with heads shaped so that liquid would condense inside the head and drain out through a collecting tube. Later, during the Middle Ages, Muslim alchemists, who also employed the *ambix*, added the Arabic article *al-* to its name, hence the term “alembic” for a still (Forbes 1948). When larger amounts of alcohol began to be distilled in the fifteenth and sixteenth centuries, the *ambix* was improved, giving rise to several types of stills.

A common one, the pot still, dates from the sixteenth century. The fermented beverage is boiled in a large pot having a curved top that allows the steam to pass to a coiled cooling pipe called the “worm,” in which the vaporized alcohol condenses into a liquid. The distillate then flows from the worm into a receiving vessel. Pot stills, like those used in the sixteenth century, were later carried to the Americas and to other European colonies and remain in use in traditional distilleries. In Europe and the United States, pot stills are employed for the production of such beverages as French brandy, Italian grappa, and Scotch whiskey (Maresca 1992). The liquids produced by a

pot still at the beginning (termed “foreshots” by American distillers) and end (“aftershots”) of a particular distillation are undrinkable or, at least, foul tasting because of the fusel oils that they contain and are consequently either redistilled or discarded. The liquid produced in the middle of the distillation process is the valuable fraction, and proper separation of the liquor from the by-products requires both experience and skill.

The early nineteenth century saw the invention, by Aeneas Coffey, of a still that permitted more or less continuous distillation (Forbes 1948). This device consisted of two hollow metal columns through which vaporized alcohol rose to condense on metal plates. One column served for distillation and the other for rectification – a process of adjusting the potency of an alcoholic beverage, often by redistillation. But because distillation was continuous, the Coffey still did not permit the separation of the bad-tasting beginning and end by-products from the vital middle of the run. All of it tasted the same, being of uniform but lesser quality. Today, such mechanical stills can process thousands of gallons of liquid at a time, but pot stills are said by many to produce better-quality beverages.

Redistillation is sometimes called “rectification,” but this is an ambiguous term, which, legally, can also mean blending one distilled beverage, or “spirit,” with other spirits or flavorings. In many cases, tasteless neutral spirit becomes an esteemed beverage with a little help from the rectifying process; examples include some Scandinavian akvavits and Polish vodkas (Grossman 1989). Spirits are also blended with water to make them drinkable.

Alcohol and Distillation – 5500 B.C. to A.D. 1500

Alcoholic beverages were made as long ago as the sixth millennium B.C., as has been documented by the discovery of wine remains in a container at Çatal Hüyük, an archaeological site on the Konya Plain in Turkey (Mellaart 1967). It is probable, however, that wine was produced from dates and figs even earlier (Tannahill 1988). Beer (or, more properly, ale) was in use in ancient Sumer by at least 2500 B.C. and at about that time, or soon afterward, in Egypt (Lichine 1981).

As populations became more crowded, with consequent pollution of water supplies, nearly everyone drank ale or wine. The Romans scorned the ale of the Germanic tribes but made wine a regular part of their own daily regime (Pliny 1940). As a rule, it was mixed with hot water, spices, and perhaps honey (Lichine 1981). People who drank undiluted wine were thought to be depressed or alcoholic (or both), which suggests that the Romans probably would not have been interested in diverting their distillation techniques from alchemy to the production of beverages stronger than wine, even had they thought of it (Tannahill 1988).

Nonetheless, some investigators have sought the origins of distilled beverages in ancient Rome. Pliny the Elder (Gaius Plinius Secundus) mentioned a "wine" that had to be diluted several times before it could (or should) be drunk (Pliny 1940). But this reference is to a wine kept in ceramic vessels and allowed to evaporate over many years, until the result was a thick sludge (Lichine 1981). Pliny also wrote of Falernian wine that could be ignited, that would keep for a decade, and that was strong in flavor. But at that time, wine was often boiled down into a kind of jelly to season foods and make drinks, and it was already known that the vapor from hot wine could be ignited. Thus, in the absence of archaeological evidence of "worm" coils or other kinds of still-head cooling devices, it appears that the manufacture of distilled spirits, at least on any large scale, remained in the future (Needham 1984).

The Greek alchemists, whose stills are portrayed in drawings from the Hellenistic period, may have preceded the Romans in producing small amounts of alcohol. Alchemy, which originated in Egypt and Persia and was then practiced by the Greeks and the Arabs, was a quasi-magical process by which the alchemist sought the "essence" of matter to perfect it in accordance with mystical laws (Needham 1984). Arab alchemists produced a cosmetic eye makeup, the name of which – *kohl* or *kubl* – conveyed the notion of something fine and subtle emerging from a process of distillation, and it was from *al-kohl* (or *al-kubl*), and various subsequent renderings like the Portuguese *álcool*, that the English word "alcohol" was derived (Forbes 1948).

Moreover, despite the Islamic prohibition against alcohol consumption, the Arabs are credited, at least in legend, with the spread of liquor to Europe. One well-known account, set in the early fifth century, tells the story of an Irish monk – who later became Saint Patrick – spreading the gospel in the Near East. There, he learned about stills and, on his return to Ireland, brought one with him (McGuire 1993). A more modern version of the tale has the Crusaders learning about alcohol and distillation from the Arabs and bringing home to Europe both taste and technique. Whether there is truth in either story, the historical record of these centuries does not indicate any widespread use of distilled spirits that would also have triggered their widespread misuse – a misuse that certainly would have rated mention in the literature (Schivelbusch 1992).

Distillation, then, continued to remain the property of the alchemists, who in the Middle Ages were distilling water hundreds of times over, reducing it to a residue of mineral salts that was said to be its "essence" (Forbes 1948). About A.D. 800, the Arab scholar Jabir ibn Hayyan invented a much-improved still, and during the centuries from 800 to 1000, Arab alchemists are said to have distilled wine, with its resulting "essence" employed in still further alchemi-

cal experimentation (Toussaint-Samat 1987). This may have been the first time that brandy was made. But with alcohol consumption forbidden by the Islamic religion, there was little incentive to distill such beverages in any quantity (Forbes 1948).

Meanwhile, in China, rice wine – *chiu* (actually an ale) – had long been produced and was as popular there as were grape wines in the West; Chinese ales made from millet also had been brewed for some time (Simoons 1991). However, the fermentation procedure used to make rice wine generally resulted in a stronger beverage than Western ales and beers (Chang 1977).

Like Greece and the Arab world, China also had its alchemists, but Chinese stills were different from those used in the West, having a tube on the side to drain the distillate and allow more of it to be produced. Although the date when this type of still originated is unknown (Needham 1984), fourth-century documents mention a "wine" from the "western regions" that kept for a long time and was extremely strong. This beverage might have been produced by Chinese stills; however, noted sinologist Joseph Needham (1984) has suggested that it came not from Chinese stills but from remote regions of central Asia, where a technique of concentrating alcohol by freezing it had been invented. Although Needham believes that such a technique was a precursor to distillation, he acknowledges that the evidence is confusing and that the references that describe this particular beverage are ambiguous. Moreover, one authority has suggested that, given the paucity of references to distilled spirits in Chinese history (until relatively recently), even if liquor were known, its use could hardly have been common (Simoons 1991).

By the fourteenth century, however, this may no longer have been the case. The *Ying-shih ssu-chi*, a medical work dated to 1368, draws attention to the dangers of overindulgence in distilled alcoholic beverages, and Li Shih Chen (writing in the sixteenth century) stated that liquor had reached China only with the Yuan (Mongol) dynasty of the thirteenth and fourteenth centuries (Chang 1977). Chen mentioned such beverages as "fire wine" and "burnt wine," equating the latter with arrack or distilled palm wine. It is interesting to note that this chronology of alcohol use in China parallels that of brandy, and perhaps even whiskey, in Europe (Simoons 1991).

From about 1000 to 1500, alchemists in Europe repeatedly distilled wine (adding salts to absorb the water portion of the liquid) to produce a distillate that would burn. That the eleventh-century Italian alchemist and scholar Michael Salernus had successfully produced alcohol, for example, is indicated by the following statement: "A mixture of pure and very strong wine with three parts salt distilled in the usual vessel, produces a liquid which will flame up when set on fire but which leaves other substances unburnt" (McCusker 1989: 85). (In Salernus's original

manuscript, the words underlined here are written in cipher to prevent others from learning his procedures and formulas.)

Salernus was not alone in reporting such results, and collectively, the alchemists believed that they had extracted the “essence” or “spirit” of wine and that repeated distillations resulted in *aqua vitae* – the “water of life” – which, in this case, was a kind of brandy that until about 1400 was used mostly as a medicine (Braudel 1973). Because the processes of aging and the separation of the different fractions of the distillate were unknown, this liquor would have been harsh, and even harsher if bad wine had been used for the distillation.

Albertus Magnus was another distiller and alchemist who wrote about the virtues of this substance. Its potency seemed to recommend it as a treatment for a variety of illnesses. Indeed, the distillate was acclaimed as the “quintessence,” a union of all the elements, even the key to everlasting life (Schivelbusch 1992).

The first real brandy that was not thought of as medicine is said to have been distilled in 1300 by Arnaldus de Villa Nova, a professor at the medical school of Montpellier. He said: “We call it *aqua vitae*, and this name is really suitable, since it is really a water of immortality” (Christian 1990: 25). On the other hand, both the Irish and the Scots claim to have produced liquor from grain (in contrast to brandy from wine) since the beginning of the last millennium; the Scots called it *uisge beatha* (pronounced whisky-baw), and the Irish called it *uisce beatha*. Both meant “water of life,” and the English term “whiskey” derived from them.

The precise dates when Irish and Scotch whiskeys originated may never be known. But we do know that, in the aftermath of Henry II’s invasion of Ireland (A.D. 1171), Irish “wine” was taxed. The reference to “wine” could mean honey mead, or ale, but it might also mean whiskey. Indeed, the Old Bushmills brand claims its origin from this date, suggesting that even if the St. Patrick tale is a bit fanciful, whiskey from the Emerald Isle may well have predated brandy in Europe, at least for recreational purposes (McGuire 1993). Certainly the archaeological discovery of a worm cooler and alembic pots in Ulster suggests that at least some distillation had taken place in Ireland by the late Middle Ages (McGuire 1993). However, the size of the vessels indicates domestic distillation on a small scale (E. C. 1859).

In the fifteenth century, better methods of cooling a still’s head were developed, and these allowed increased production of distilled beverages (Forbes 1948). The technology spread quickly across Europe, and practically every country developed its own national distilled spirit. Those countries also soon developed laws to tax, restrict, and sometimes even ban such spirits, because by the sixteenth century, drunkenness had become a serious social problem.

People who had previously drunk beer as if it were water were discovering that they could not drink liquor as if it were beer (Schivelbusch 1992).

Centuries earlier, Paracelsus (Philippus Aureolus) had employed the Arabic term *alcool vini* to describe spirits. But it was not until 1730, when the Dutch physician Herman Boerhave used the word alcohol to mean distilled spirits, that it became commonly understood that ale, wine, and distilled beverages all owed their mood-altering capabilities to this chemical (Forbes 1948). Yet because “*aqua vitae*” and other such appellations continued to be used, it is difficult to discover exactly what kinds of spirits were actually being produced. Indeed, even Scotch whiskey was called *aqua vitae* (Jackson 1988).

From this point forward, however, the historical picture is sufficiently clear to permit treatment of the individual liquors, and we will attempt to do this in some semblance of chronological order. But before we begin, a word or two is needed about the strength, or “proof,” of an alcoholic beverage. The term *proof* originated in connection with the early use of gunpowder in war: “Proof,” or good, armor was that which proved resistant to a gunshot. The word entered alcohol terminology as a means of identifying the quality of rum and brandy. “Proof” beverages were of the approved strength – half spirit and half water (McCusker 1989). Their purity could be measured by weighing or by setting the spirit alight. Later, the term came to mean twice the percentage of alcohol in the drink. In the twentieth century, neutral spirit leaving the mechanical still is 180 proof, or 90 percent alcohol (Grossman 1989).

Distilled Spirits in the West

Brandy

First called “brandy wine” (from the Dutch *brandewijn*), brandy means “to burn” or “burnt” in Dutch as well as in other languages, such as the German *Brand* and the Middle English “brand.” Brandy is more expensive to make than grain spirits because it must be distilled from fruit and, in the case of cognac, from wine (Ray 1974). As noted, brandy first emerged as a medicine in the eleventh century and only later became popular as a beverage. In Nuremberg, it had apparently become a bit too popular by 1450; in that year, a law was passed to ban the drinking of *aquavit*, which we are told was brandy wine (Forbes 1948). Because in northern Europe beers and ales had long been made by women, during the fifteenth and sixteenth centuries many distilleries were located in private homes. Nonetheless, by 1500, brandy production was subject to taxation in many principalities (Braudel 1973).

Gradually, the French wine country became the center of the brandy industry. Louis XII granted the first license to manufacture brandy in 1514, and the product was employed (among other uses) to fortify

wines, which strengthened them at the same time that it stopped fermentation (Braudel 1973). A less expensive brandy for the poor was made from leftover grape skins and seeds. It was and is called *grappa* in Italy, *marc* in France, *aguardente* in Portugal, and *aguardiente* in Spain (Maresca 1992).

A countermovement that pushed up the price was the aging of brandy, a procedure begun about the middle of the eighteenth century (Ray 1974). This process made for a decidedly smoother drink and certainly helped to promote what has been termed the "cult" of brandy, just as the aging of Scotch and Irish whiskeys in oak barrels established a whiskey "cult" (Morrice 1983). Then, as now, the most famous brandies came from the Charente region of France and were named for the town of Cognac. Twice distilled in pot stills from the white wines of the Charente, cognac is blended according to certain formulas and aged for a minimum of three years in oaken barrels. The drink can continue to improve through aging for as many as 50 more years, but that is the effective limit of the process; thus, the legend of fine brandy surviving from the time of Napoleon is a myth. Experts agree that even if such brandy existed, it would be undrinkable (Ray 1974).

Cognac leaves the still at 140 proof, but its potency is reduced by aging because alcohol evaporates through the porous material of the cask, whereas water does not. In addition, sufficient water is added so that cognac is shipped at 80 to 86 proof. The age and quality of a cognac are indicated by a confusing array of letters, stars, and symbols on the label. Armagnac, a well-known brandy from another region of France although much like cognac, is said to retain the flavor of the wine to a greater degree, and its method of distillation is different (Ray 1974).

By the late nineteenth century, the brandy industry was in near ruin because of a vine blight caused by the American vine louse, *Phylloxera*, that destroyed virtually all the vines in Europe (Lichine 1981). A visitor to the Continent at the end of the nineteenth century reported that the "brandy" available was not brandy at all, but rather an ersatz mixture concocted with grain spirits and plums (Spencer 1899). However, the European wine industry, along with the manufacture of brandy, was saved by the introduction of resistant rootstocks of muscadine and scuppernong grapes, native to North America, onto which the famous vines of Europe were grafted (Lichine 1981).

The term, brandy, refers not only to drinks made from grapes and wine but also to a wide variety of beverages made from other fruits and even from honey. Some are true brandies, made entirely from fruit, but many are grain spirits that are merely flavored with fruit. One fruit brandy is kirsch (*Kirschwasser*), a fiery cherry brandy of Switzerland. *Poire* brandy is made from pears, and *framboise* from raspberries. Hungary produces *Barack Palinka*, perhaps the most famous of apricot brandies. Another

important beverage is slivovitz, the plum brandy of the Balkans, sold at 70 and 87 proof. Most of these brandies are bottled as uncolored spirits to preserve the fruit bouquet. In addition, there are countless brandies made from apples in Europe and wherever else apples are grown. Examples are calvados, made in Normandy from hard cider, and applejack, made largely in North America. Calvados is aged, whereas applejack frequently is not, and historically, applejack has been made by freezing as well as by distillation (Grossman 1977; Lichine 1981).

In the Americas, the agave plant contains a juice that is fermented into *pulque*, a beverage that dates from the Aztec period. The brandy *mescal* that is distilled from *pulque* comes in many varieties, with the most famous being tequila (Grimes 1988). In Peru, grape wine has been produced since 1566, with *pisco* the local brandy. Both Mexico and Chile also produce brandy made from grape wines (Lichine 1981).

Moving northward, in colonial New England and in Canada, hard cider was first processed into applejack by freezing, a technique reported much earlier in central Asia (Dabney 1974). In the Appalachian Mountains, whole valleys were planted with fruit trees to make apple and other brandies until the early twentieth century (Marcus and Burner 1992). Although there are references to "British brandy" in seventeenth-century texts, these actually deal with an early grain spirit related to gin; Britain produces virtually no wine or brandy. However, a few British distillers have made brandy from mead, the honey-based wine of the ancient Britons. This beverage, called "honey brandy," has never been common because mead itself is costly, and distilling it yields a very expensive liquor (Gayre 1948).

Spirits from Grain and Cane

The kinds of grain used to make spirits have usually been determined by custom (although sometimes by law, as in the case of licensed distillers) and, perhaps most importantly, by what is available. As a rule, lower-cost and lower-status grains have been employed: To make spirits from wheat would have been a waste of bread grain, and in the sixteenth century, several German states banned the manufacture of grain spirits so as better to control the price of bread (Forbes 1948). Rye has commonly been used in areas with colder climates, such as Canada. In the United States, the abundance of maize has made it the grain of choice (Inglett and Munck 1980), and barley constituted the base of the first whiskeys made in Ireland and Scotland. Many of these grains were ingredients for beers long before whiskeys were made (Jackson 1988).

Irish and Scotch whiskeys. To produce either Irish or Scotch whiskey, soaked barley is permitted to sprout and become malt, as in making beer. However, the mash for whiskey does not include hops. The Scotch malt has traditionally been dried over peat fires that

imparted a smoky flavor, whereas Irish malt was dried in ovens. After the grain is made into a mash and allowed to ferment, it is distilled in a pot still – twice to make Scotch and three times to make Irish whiskey (Jackson 1988). The spirits are then permitted to age in oaken casks (from which they acquire their color); then blending takes place. This is a procedure that employs whiskeys that have been aged for different periods of time and that are often made from different ingredients (Morrice 1983). In the case of Scotch, for example, the unblended or “pure” malt can be blended with whiskeys made from other grains, such as oat, rye, and maize.

Modern Scotch seems to date from about the middle of the eighteenth century, its production spurred by a 1745 increase in the tax on ale. Private distillation was banned in the nineteenth century (Morrice 1983). However, the distillation of whiskey was an important source of income for the Highlanders, who raised grain but encountered much difficulty in getting it to market over the rough mountain roads. Consequently, the ban gave rise to a thriving illegal industry that, in distilling the grain, put it into a much more compact (and valuable) form. Doubtless the industry became more profitable after the *Phylloxera* blight that devastated European wine grapes from the 1860s on, made brandy virtually unaffordable, and as a result, increased demand for other alcoholic beverages.

Pot stills were hidden away in the Scottish Highlands and in Ireland, where clans made illegal whiskey, as their kinsmen in America were soon to do as well. The government struck back with raids on the illegal stills; in 1834, there were 692 seizures in Scotland and 8,192 in Ireland. But in 1884 there were only 22 seizures in Scotland, and in Ireland they had fallen off to 829. Clearly, the government was winning the war. Perhaps the greater amount of government activity in Ireland had to do with the fact that in the 1880s, there were only 28 legal distilleries, in contrast to 129 in Scotland (Barnard 1969). Government licensing was profitable: In the 1880s, the government of the United Kingdom derived 40 percent of its revenue from alcohol (Christian 1990).

Scotch whiskey continues to enjoy a certain mystique. But its manufacture has changed considerably, with coal employed to dry the malt instead of peat. The grain spirit most used to blend Scotch is now made from American maize, and Scotland imports all of its maize and much of its barley (Morrice 1983). Somehow, too, much of the romance has disappeared.

Gin. Gin, on the other hand, never had all that much romance attached to it, especially after William Hogarth depicted the excessive drinking of Londoners in his 1751 engraving, “Gin Lane” (Schivelbusch 1992). Less than a century before, the consumption of spirits had been uncommon in England (Austin 1985); most people drank ale or beer, and the wealthy, although

inclined to a little French brandy, mostly prized imported wines, especially those that were fortified, such as port and Madeira from Portugal and its islands and sherry from Spain (Lichine 1981). Indeed, the dramas and diaries of the Restoration period indicate that these wines were preferred by many over even the “great growths” of France.

The individual said to have pushed Britain onto the path toward “Gin Lane” is William of Orange, Stadtholder of Holland, who ascended the English throne in 1689 as joint sovereign with Mary II. William’s homeland made gin, and to encourage its importation into England, he discouraged the sale of French brandy and taxed English beer and cider (Watney 1976). The liquor he sponsored with such determination was then called *genever* or *jenever*, the Dutch name for juniper, a medicinal plant whose berries were used to flavor gin, which was itself a distillation of hopped barley mash (Butler 1926). Also named “Hollands gin” by the English, the drink became so popular in the Netherlands that it replaced beer as the beverage of the military (Austin 1985). In fact, by 1787, there were more than 200 distilleries making gin in the Dutch Republic, and only 57 breweries (Schama 1977).

Many of the gin distilleries had doubtless come into existence to supply the English market that William had engineered. In 1700, some half-million gallons were exported to England (George 1965). But it did not take the English long to begin making their own gin, which, they found, was a good way to use up surplus grain. After all, gin could be made from practically any grain, or even from molasses; its distinctiveness lay in the juniper flavoring (Watney 1976). In 1714, 2 million gallons were produced, and by 1737, English distillers were making 5.4 million gallons annually – nearly a gallon for every man, woman, and child in the population (George 1965).

As early as 1725, London alone held 6,187 gin shops, where people unaccustomed to strong spirits were working hard to overcome this handicap (Monckton 1966). It was called a “gin epidemic”: Tavern keepers offered to make one “drunk for a penny, dead drunk for two pence, and straw for nothing” (Schivelbusch 1992: 156), meaning that drunkards were thrown into cellars strewn with straw to sleep off their intoxication. In London during a week in 1750, some 56,000 bushels of wheat were used to produce gin, not quite as many as the 72,000 bushels devoted to breadmaking, but close (Monckton 1966).

At this time, however, gin consumption was about to decline. The Gin Act, passed by Parliament in 1736, prohibited the public sale of gin in London and raised taxes on spirits (George 1965). At first, this statute had little effect on gin consumption: Bootleggers, some employing subterfuges such as the labeling of gin as medicine, helped perpetuate the “epidemic” (Watney 1976). Then in 1751, Parliament slapped a very high tax on gin and tightened restrictions on its

sale. This effort, coupled with a sudden rise in grain prices, pushed the price of gin beyond the reach of the poor, who switched back to ale and beer, or to coffee and tea, beverages that were becoming increasingly popular (Mintz 1985). By 1782, annual gin production had fallen to 4 million gallons, and a century after that, Britons were consuming annually only about 1.25 gallons per capita of all spirits combined, although per capita beer and ale consumption was 34 gallons (Johnston 1977).

The Beefeaters distillery was established in 1822, and during the nineteenth century, gin was flavored with angelica, cassia, cardamom, coriander, and ginger, as well as juniper berries (Lichine 1981). The martini cocktail is said to have been invented in 1863 by San Francisco bartender Jerry Thomas (who is also credited with inventing the Tom and Jerry). He allegedly named the drink after the town of Martinez upon learning that this was the destination of a departing guest. However, the use of gin in martinis and "gibsons" (the latter garnished with a pearl onion instead of an olive) has chiefly been a twentieth-century phenomenon. In addition, gin is used in many mixed drinks because of its somewhat neutral taste (Lanza 1995). Modern gin is made in industrial stills and is between 80 and 97 proof (Grossman 1977). The most common type of gin is "London dry" made in both Great Britain and the United States. English gins are 94 proof; American versions are between 80 and 94 proof. Dutch gin survives today as a strongly flavored beverage distilled at a lower proof and drunk neat because of its taste (Grossman 1989).

Vodka. Vodka's distillation apparently dates from the sixteenth century, as the Russians followed the example of others in northern Europe and distilled rye – their most abundant grain – into vodka, using the pot still (Pokhlebkin 1992). As with other liquors, the distillation of vodka produced much leftover mash that could be fed to livestock, making it possible to maintain more animals over harsh winters (Christian 1990). Later, in the nineteenth century, potatoes were employed for distillation, but the thick potato mash required special stills. It has been mostly in the twentieth century that maize and wheat have found their way into Russian and Polish stills.

Vodka is thought of by some as a relative of gin, and many of its uses are the same. On the other hand, the Russians and the Poles point out that vodka "is the only beverage that goes with herring, a fish that makes beer taste insipid and wine metallic" (Tousaint-Samat 1987). Vodka means "little water," and the drink has been valued for its colorlessness and purity. It was also much valued by the Russian government for producing revenues (Christian 1990). Prior to the creation of vodka, the bulk of the country's alcohol production was in the form of beers and meads made locally by the peasantry and, thus, not taxable. But vodka was taxable and, in 1861, provided 45 percent

of the revenue of the Russian state. Throughout the nineteenth century, more than 70 percent of expenditures for the Russian army were paid for by vodka taxes (Christian 1990). At the turn of the twentieth century, Russia established its Kristall vodka monopoly (under the control of the ministry of finance), which produced Stolichnaya vodka and some 70 other brands of liquor (Lichine 1981). The monopoly was suspended from 1914 to 1921, as the various Russian governments during World War I, the Bolshevik revolution, and the Russian civil war attempted prohibition (Pokhlebkin 1992), and was finally abandoned following another wave of government-encouraged temperance begun in the 1980s. Vodka is normally between 80 and 100 proof (Grossman 1977).

Grain spirits similar to gin and vodka are common across northern Europe. One example is Scandinavia's akvavit (Lichine 1981) – a name that is variously spelled. This liquor is distilled from a grain or potato base as a neutral spirit, then redistilled with flavoring. Like gin, akvavit is not aged (Grossman 1977). German and Scandinavian schnapps are usually made from fruit or herbs; Dutch *genever* is flavored with sweet-smelling herbs and is also called schnapps in Europe (Lichine 1981).

Rum. In North America, grain and cane spirits were favored over ale, beer, and wine from the beginning of European settlement (Rorabaugh 1979). English ale would not ferment properly in America's climate; cold winters froze it and hot summers caused the top-fermenting yeast to spoil. (Modern American beer is made with German lager yeasts brought by immigrants in the 1840s. These live at the bottom of the vat and can thus withstand extremes of temperature.) Nor did European wine grapes do well on the Eastern seaboard. Some distillation of grain spirits took place and, though never completely abandoned, was put aside as it became evident that the most expedient course was to ship grain to the sugar islands of the West Indies and to make into rum the molasses sent northward in return.

Rum became the favorite American alcoholic beverage until the Revolution. It is produced by distilling the fermented alcohol made from sugarcane juice (sometimes called "dunder" or "burned ale"). The best rums are distilled twice and, in some cases, aged. Rum was made in Barbados in the 1630s and received its name in 1651, when traveler Thomas Ligon remarked that the island's inhabitants were fond of "Rumbullion alias Kill Devil, and this is made of sugar canes distilled, a hot, hellish and terrible liquor" (Ritchie 1981: 116). Both "rumbullion" and "rumbustion," two early dialect names, may have referred to the violence that the drink was said to engender. Rum was shipped to England in the 1660s, but a taste for it there was slow to develop. In 1698, the English imported only 207 gallons. By contrast, the West India colonists drank so much of it that a visitor said they had "bodies like

Egyptian Mummies" (McCusker 1989). Slaves often drank a colorless raw alcohol made from sugarcane juice, called *clairin* in Haiti, *aguardiente* in Mexico, and *cachaça* in Brazil (Lichine 1981). Distilled only once, and devoid of the flavoring and aging that makes a good rum, these harsher beverages actually were rums, although not called by that name; the term "raw rum," however, is sometimes used to refer to *clairin*.

The demand for rum was great in Africa, where it was traded for slaves to be carried to the West Indies to make more sugar. Much rum was also issued to the slaves of the Caribbean, where it subsequently was incorporated into Afro-Caribbean religious and magical rituals (Lichine 1981). The spirit also became part of the medical lore of the region for a few brief decades spanning the end of the seventeenth century and the first half or so of the eighteenth century, when slaves, soldiers, and sailors, who could afford (or were given) only the cheapest rum, came down with the "dry bellyache." The almost unbearable cramps characteristic of this disease were, in fact, symptoms of lead poisoning – the lead having entered the rum from lead fittings in the stills (Handler et al. 1987). Although diminished in frequency, the dry bellyache persisted as late as the Victorian period (Spencer 1899).

Demand for rum in the North American colonies was intense, requiring distilleries in the sugar islands and in New England, South Carolina, Pennsylvania, and the Canadian Maritime Provinces to satisfy it (McCusker 1989). William Penn's thirsty colonists drank rum, and by the 1720s, Marylanders were managing a per capita consumption of at least 2 gallons annually. Rum was relatively cheap: In 1738, rum from any of Boston's 8 distilleries (they numbered 63 by 1750) cost 2 shillings a gallon. In Georgia, a day's wages could keep a laborer drunk for a week. By 1770, the 118 recorded colonial distilleries were producing an annual total of 4,807,000 gallons of rum, an average of 41,000 gallons each (McCusker 1989).

The West Indies sugar industry inexorably became the foundation of British mercantilism that in 1733 led to the passage of the Molasses Act by Parliament. This law heavily taxed molasses and rum shipped to North America from the non-British Caribbean. It drove up the price of inexpensive molasses from French St. Domingue which caused a substantial increase in rum prices. As a consequence the Molasses Act began decades of rum and molasses smuggling that ended only with the American Revolution. In addition to rum being their favorite drink, much of the prosperity of the North American colonists depended on the molasses trade (McCusker 1989).

In 1759, British grain crops failed, leading to an increase in grain prices throughout the following decade, with the result that the English (especially

those of the upper classes) began to develop a late appreciation for rum. Indeed, by 1775, England was importing 2 million gallons of rum each year. Much of this was used to make rum punch, although many thought this drink was inferior to Madeira punch, made from the fortified wine of Madeira (Tannahill 1988).

Rum became even more profitable for British planters after the Molasses Act. In 1798 in Jamaica, for example, planters who made £3,000 net profit from sugar could count on an additional £1,300 from rum (Deerr 1950). By this time, the French had been all but removed from the sugar and rum production of the West Indies because of the revolution of St. Domingue and the emergence of an independent Haiti. Yet Britain was about to be more or less eliminated as a major sugar producer as well. In 1807, the British abolished the slave trade and, in 1833, slavery itself. At this point, Brazil and Cuba became the world's most important sources of sugar, with Cuba outdistancing Brazil as the nineteenth century progressed. The Bacardi Rum Company, founded in Cuba in 1862, became the producer, in the twentieth century, of the most popular brand of rum in the world.

Cuban rum tends to be light, as compared with the darker rums from Jamaica and Barbados, the latter often double distilled and aged (Deerr 1950). Haitian rum is midway between these extremes. There is also *arak* rum from Java, made from sugarcane juice fermented with rice and yeast (Grossman 1977). The yeast strains used to ferment sugarcane juice are said to be responsible for the variations in flavor among different kinds of rum (Lichine 1981). However, as sugarcane always ferments within a short time after being cut, it seems clear that wild yeast is adequate to the task of fermenting the cane juice.

Rum has a strong association with the sea and sailors. Those of the United States, Canada, and Britain were huge consumers of rum, and "grog" rations in the latter two navies continued into the 1970s. The term "grog" comes, allegedly, from the eighteenth-century British Admiral Edward Vernon, who wore a grogham (heavy wool) cloak, and to reduce drunkenness among his crews, ordered their rum diluted with water, resulting in the famous rum-based drink named after his outer garment (Tannahill 1988).

Bourbon and other New World whiskeys. Although rum was the drink of colonial Americans, independence from Great Britain severed much of the sugar connection, forcing a greater reliance on local resources. The result was a change to corn whiskey or "bourbon" (as it came to be called), made originally in Bourbon County, Kentucky. By 1800, Americans were drinking some 3.5 gallons of liquor – mostly whiskeys – per capita each year, which seems to represent a substantial increase in consumption over that of rum in previous decades.

Not that Americans had ever relied solely on rum. Grain whiskeys made from rye and corn had been produced on the eastern seaboard practically from the beginning of the colonies, and whiskey was also imported. In fact, each family that arrived with the Winthrop fleet of 1629 was required to carry *aqua vitae* to the Massachusetts Bay Colony, presumably as a medicine (Cressy 1995).

During the eighteenth century, Scottish and Irish settlers brought with them the techniques of whiskey making. Many settled in the southern colonies as farmers, and following the American Revolution, distillation was one way in which farmers might add to their yearly income. Although such activities were pursued virtually everywhere in the new United States, a special enthusiasm for operating stills developed in the back country of the Blue Ridge and Great Smoky Mountains, where (as in Scotland and Ireland) rugged terrain hindered the transportation of grain to market. Although an acre of land might produce 40 to 60 bushels of corn, a horse could carry only four bushels at one time. The same horse, however, could carry 16 gallons of whiskey, which was the equivalent of 24 bushels of grain. Clearly, for the mountain people, distillation represented the most efficient method of preparing corn for sale. In the 1830s, consumption of distilled spirits in America reached a peak of more than 5 gallons per capita annually (Rorabaugh 1979); by this time, whiskey made from corn accounted for most of that amount (Dabney 1974).

Private distillation, though illegal, continued even after roads were constructed through the mountain areas, partly because of tradition, partly because of remoteness (stills were easy to conceal), but mostly because it was increasingly profitable, as taxes on liquor continued to be raised. Thus, although illegal whiskey has been made as far afield as Alaska (where it was called "hooch" after the Hoochinoo, a tribe that made a distilled drink), 60 percent of all arrests for the offense have been made in the mountains of the South (Carr 1972).

Corn, however, was no longer always the primary ingredient in illicit distillation. Sugar was sometimes used to speed up fermentation, resulting in a kind of rum that was, nevertheless, still called whiskey. Sugar whiskey was notorious for the hangovers and headaches it produced. In addition, illegal whiskey was sometimes adulterated with everything from manure to lye (to speed fermentation). It was often distilled in soldered tanks that put lead salts into the liquid, and it was condensed in old automobile radiators. Clearly, illegal whiskey could be dangerous: It killed some people, blinded more, and helped destroy the brains, livers, and stomachs of many others (Dabney 1974).

Today, Kentucky is the center of whiskey production in the United States, although some is made in Tennessee and Virginia (Jackson 1988). Starchy vari-

eties of corn are best for whiskey making. They are wet or dry milled, then fermented before distillation. The two major yeasting processes are sweet and sour mash. Sweet mash whiskey comes from new yeast, whereas yeast from a previous fermentation constitutes one-fourth of the yeast added to grain to produce sour mash whiskey. Sour mash is easier to make, and sweet mash tends to go bad (Grossman 1977). In the United States, a significant portion of the corn crop goes into brewing and distillation (Inglett and Munck 1980). Obviously, bourbon remains an American favorite. It is made from at least 51 percent corn mash distilled at 160 proof, diluted to 80 to 100 proof, and matured for no less than three years (Lichine 1981).

Canada, which produced its share of rum in the Maritime Provinces, turned to whiskey made from rye, now mostly called Canadian whiskey (Jackson 1988; Morrison and Moreira 1988). Straight whiskeys are the products of distillation with nothing added, whereas blended whiskeys are mixtures of whiskey from different distilleries or different years. These two kinds, straight and blended combined, make up about half of the American whiskey market. The remainder is "light" whiskey, which is less flavorful, always blended, and reduced in proof. Like Canadian whiskey, it is often made from rye.

Liqueurs. Liqueurs are grain or cane spirits that have been flavored, are usually sweet, and are normally enjoyed after dinner. Their origin lies in the early years of alcohol distillation, when sugar and flavorings were employed to mask the bad taste of raw alcohol, and until the sixteenth century or so, they were mostly regarded as medicines. Early in that century, however, this began to change. In 1532, Michael Savonarola, a Florentine physician, authored *The Art of Making Waters*, a book of recipes and instructions for producing liqueurs. By the following century, Italy had a flourishing liqueur industry (Austin 1985).

Many of the old liqueur recipes, with "secret" mixtures of flavorings, originated in monasteries. The most famous liqueurs made from such recipes are Chartreuse (from the Chartreuse monastery at Paris) and Benedictine (Grossman 1977). Perhaps the most notorious liqueur is no longer made. This was absinthe, flavored with hallucinogenic wormwood (*Artemisia absinthium*) that caused a variety of physical and mental symptoms, and even death (Conrad 1988). It was produced by Henri-Louis Pernod (Lanier 1995). France finally outlawed absinthe in 1915, and now it is illegal virtually everywhere (Lichine 1981). Liqueurs are flavored with ingredients as mundane as coffee (Kahlua, made in Mexico) and as exotic as the rare green oranges of the Caribbean (Curaçao, from the island of that name). Other famous coffee liqueurs are Tia Maria from Jamaica and Pasha from Turkey (Grossman 1977).

Distilled Spirits in the East

Spirits are not as popular in the East as in the West, despite remarkable historical exceptions, such as the hard-drinking Mongols, who have even been credited with the discovery of distillation. The lukewarm eastern attitude toward liquor has existed in part because of the popularity of tea, in part because opium, betel nut, and other stimulants have historically substituted for strong drink, and in part because of a widespread devotion to beer. Another reason may be that the liver of Asian peoples tend to be low in aldehyde dehydrogenase isozyme (ALDH), which helps to metabolize alcohol. This condition can cause the face to flush with the ingestion of even very moderate amounts of alcohol – a telltale signal among many who live by Confucian rules of conduct that frown on intoxication. In nineteenth-century imperial China, when strong drink was served, a gentleman not wishing to offend his host might pay someone else to drink it for him (Chang 1977).

Nonetheless, several grain alcohols are produced in the East. In China, *mao tai* – a whiskey made from millet and wheat – is the best known, and has been known to the world since it was served to U.S. President Richard Nixon during his 1972 visit there (Simoons 1991). *San-shao*, meaning “three times burnt,” is another Chinese liquor. It is made from sorghum and distilled three times, but is not to be confused with *shao-chiu*, which is distilled from grain ale (Simoons 1991). In Japan, the latter is called *shochu* and competes with local whiskey that is modeled on Scotch (Jackson 1988).

In Mongolia and Siberia, the drink best known from lore and literature is *kumyss* (or *kumiss*), an alcoholic beverage (fermented and sometimes distilled) made from mare’s milk. In about the seventh century, the Mongols introduced it in China, where it was popular for a time. Later, however, it acquired a reputation as a drink of barbarians and was so regarded after the Mongols conquered China in the thirteenth century (Chang 1977). A high level of lactose intolerance among the Chinese may also help to explain the disdain for *kumyss*, as well as for other milk products. The distilled form of the drink is sometimes called *arak*, a common Asian name for all kinds of liquor (Lichine 1981).

In Southeast Asia, “toddy,” a distilled palm wine made from fermented sap, appears in many varieties that may actually be thought of as brandies. Historically, toddies have been made in primitive stills, although a few are commercially produced (Lichine 1981). However, toddy, and other distilled beverages that exist in Oceania, are poorly documented. The latter have resulted from the distillation by Europeans of local alcoholic beverages; one example is the *okelabao* of Hawaii. This drink, made from *ti* roots fermented in the bilge of a canoe and distilled in a ship’s cookpot, has yet to find a market outside the islands (Grossman 1977).

Distilled Spirits and Human Health

Chemically, any distilled beverage consists largely of ethyl alcohol. Other components include esters, fusel oils (isobutyl and amyl alcohols) and, of course, the ingredients added to flavor and color the beverage. Unflavored grain spirit is merely alcohol and water. It is the impurities and additives that largely contribute to hangovers (Dabney 1974). As mentioned, early distilled alcohol was used only as a medicine because distillers did not yet know how to separate the unpleasant beginning and end fractions of the distilled liquid from the middle, and because aging, which mellows the product, was unknown. The effect of the early medicinal wine brandies must have been strong indeed.

Nutritionally, distilled beverages are high in calories but contain little in the way of other nutrients. Because each gram of 86 proof alcohol imparts 7 calories, the average drink bristles with 106 calories, in addition to any calories in the mix (Robertson et al. 1986). Some researchers, however, contend that the calories in alcohol have, in the past, served as an important source of energy for the poor (Braudel 1973). One study suggests that in France during the 1780s, 10 percent of an individual’s caloric intake was supplied by alcohol (Austin 1985). The same was probably the case for slaves in the Caribbean: Jamaican rum yields twice the calories of a similar measure of molasses. A counter-argument, of course, is that alcohol was allowed to replace more nutritional foods. John McCusker (1989) has noted that alcohol calories are more quickly absorbed than those from other sources; he credits the ability of early Americans to consume such large amounts of rum and whiskey to this propensity.

Yet, a large amount of any kind of alcohol can be nutritionally disastrous because it destroys vitamin C and the B-complex vitamins. Indeed, about the only remaining cases of frankly nutritional diseases (such as pellagra) found in the developed world are among alcoholics. In addition, although it appears that moderate alcohol consumption can help prevent heart disease, large amounts can help cause it. Alcohol is also suspected of being a factor in the etiology of some cancers and is known to be a culprit in causing much liver damage, including cirrhosis.

Another unfavorable aspect of distilled spirits is the social and physical harm they have historically brought to peoples unaccustomed to them (Mancall 1995). In America, Asia, Africa, and Australia, rum and whiskey became instrumental, in the hands first of European traders and then of European imperialists, in destroying aboriginal life (Miller 1985). It is significant that in Mexico and Peru, where some alcoholic beverages existed at the time of European contact, the aboriginal peoples and their traditions have fared much better than those in places like Australia, where alcohol had been unknown.

Clearly, distilled spirits have had a tremendous impact on human history and health. In a relatively few centuries, their manufacture has moved from the quasi-magical procedure of the alchemists to a global industry that undergirds the economies of entire regions. But from the "gin epidemic" of England to the endemic drunkenness of the Australian aborigines, spirits have also caused such misery that practically every society in the world has laws and customs to regulate their consumption, and many states have tried to outlaw them. That such attempts have been largely unsuccessful demonstrates the existence of a worldwide, collective opinion about the pleasures and profits provided by alcohol, which outweighs the harm it continues to cause.

James Comer

Bibliography

- Austin, Gregory. 1985. *Alcohol in Western society from antiquity to 1800: A chronological history*. Santa Barbara, Calif.
- Barnard, Alfred. 1969. *The whisky distilleries of the United Kingdom*. Newton Abbot, England.
- Barty-King, Hugh, and Anton Massell. 1983. *Rum*. London.
- Benes, Peter, ed. 1985. *Foodways in the Northeast*. Boston.
- Bergeron, Victor. 1946. *Trader Vic's book of food and drink*. Garden City, N.Y.
- Braudel, Fernand. 1973. *Capitalism and material life, 1400-1800*. New York.
- Bruce-Mitford, Rupert. 1975. *The Sutton Hoo ship burial*. London.
- Butler, Frank. 1926. *Wine and the winelands of the world, with some account of places visited*. London.
- Carr, Jess. 1972. *The second oldest profession; an informal history of moonshining in America*. Englewood Cliffs, N.J.
- Chang, K. C., ed. 1977. *Food in Chinese culture*. New Haven, Conn.
- Christian, David. 1990. *Living water*. Oxford.
- Conrad, Barnaby. 1988. *Absinthe: History in a bottle*. San Francisco.
- Cressy, David. 1995. *Coming over: Migration and communication between England and New England in the seventeenth century*. New York.
- Dabney, Joseph Earl. 1974. *Mountain spirits. A chronicle of corn whiskey from King James' Ulster Plantation to America's Appalachians and the moonshine life*. New York.
- Deerr, Noel. 1950. *The history of sugar*. London.
- Doxat, John. 1950. *The world of drinks and drinking*. New York.
- Driver, Harold. 1961. *Indians of North America*. London.
- E. C. (no other name given). 1859. On the antiquity of brewing and distilling in Ireland. *Ulster Journal of Archaeology* 7: 33.
- Forbes, R. J. 1948. *Short history of the art of distillation from the beginning up to the death of Cellier Blumenthal*. Leiden, Netherlands.
- Gayre, G. R. 1948. *Wassail! In mazers of mead*. London.
- George, Mary Dorothy. 1965. *London life in the eighteenth century*. New York.
- Grimes, William. 1988. Rio brandy. *Esquire* 110: 18.
- Grossman, Harold. 1977. *Grossman's guide to wines, beers, and spirits*. Sixth revised edition. Rev. Harriet Lembeck. New York.
- Handler, Jerome S., Arthur C. Aufderheide, and Robert S. Corruccini. 1987. Lead contact and poisoning in Barbados slaves. In *The African exchange*, ed. Kenneth F. Kiple, 140-66. London.
- Harrison, Brian. 1971. *Drink and the Victorians*. Oxford.
- Holmes, Urban Tignor. 1952. *Daily living in the twelfth century, based on the observations of Alexander Neckam in London and Paris*. Madison, Wis.
- Inglett, George, and Lars Munck, eds. 1980. *Cereals for food and beverage*. New York.
- Jackson, Michael. 1988. *The world guide to whisky*. Topsfield, Mass.
- Johnston, James. 1977. *A hundred years of eating*. Dublin.
- Kiple, Kenneth F. 1984. *The Caribbean slave: A biological history*. London.
- Kurtz, Ernest. 1979. *Not-God. A history of Alcoholics Anonymous*. Center City, Minn.
- Lanier, Doris. 1995. *Absinthe, the cocaine of the nineteenth century: A history of the hallucinogenic drug and its effect on artists and writers in Europe and the United States*. Jefferson, N.C.
- Lanza, Joseph. 1995. *The cocktail: The influence of spirits on the American psyche*. New York.
- Lichine, Alexis. 1981. *New encyclopedia of wines and spirits*. Third edition. New York.
- Mancall, Peter. 1995. *Deadly medicine: Indians and alcohol in early America*. Ithaca, N.Y.
- Marcus, Robert D., and David Burner, eds. 1992. *America firsthand: Readings in American history*. Second edition. New York.
- Maresca, Tom. 1992. Grappa glorified. *Town & Country Monthly* 146: 40-2.
- Maurer, David. 1974. *Kentucky moonshine*. Lexington, Ky.
- McCusker, John. 1989. *Rum and the American Revolution: The rum trade and the balance of payments of the thirteen continental colonies*. New York.
- McGuire, E. B. 1993. *Irish whiskey: A history of distilling, the spirit trade, and excise controls in Ireland*. New York.
- Mellaart, James. 1967. *Çatal Hüyük: A Neolithic town in Anatolia*. New York.
- Miller, James. 1985. *Koori*. London.
- Miller, Wilbur R. 1991. *Revenuers and moonshiners: Enforcing federal liquor laws in the mountain South, 1865-1900*. Chapel Hill, N.C.
- Mintz, Sidney. 1985. *Sweetness and power: The place of sugar in modern history*. London.
- Monckton, H. A. 1966. *A history of English ale and beer*. London.
- Morrice, Philip. 1983. *The Schweppes guide to Scotch*. Sherborne, England.
- Morrison, James H., and James Moreira, eds. 1988. *Tempered by rum: Rum in the history of the Maritime Provinces*. Porters Lake, Nova Scotia.
- Needham, Joseph. 1984. Spagyral discovery and invention. In *Science and civilization in China*, Vol. 5, Part 4. London.
- Nelson, Derek. 1995. *Moonshiners, bootleggers, and rum-runners*. Osceola, Wis.
- Pliny (Gaius Plinius Secundus) [Venice 1469] 1940. *Natural history*, trans. Harris Rackham, ed. W. Heinemann. London.
- Pokhlebkin, William. 1992. *A history of vodka*, trans. Renfrey Clarke. London.

- Ray, Cyril. 1974. *Cognac*. New York.
- Ritchie, Carson I. A. 1981. *Food in civilization*. New York.
- Robertson, Laurel, Carol Flinders, and Brian Ruppenthal. 1986. *The new Laurel's kitchen*. Second edition. Berkeley, Calif.
- Roesdahl, Else, and David M. Wilson, eds. 1992. *From Viking to Crusader: The Scandinavians and Europe, 800-1200*, trans. Helen Clarke et al. New York.
- Ronnenberg, Herman. 1993. *Beer and brewing in the inland Northwest*. Moscow, Idaho.
- Rorabaugh, W. J. 1979. *The alcoholic republic, an American tradition*. New York.
- Schama, Simon. 1977. *Patriots and liberators: Revolution in the Netherlands, 1780-1813*. London.
- Schivelbusch, Wolfgang. 1992. *Tastes of paradise*. New York.
- Simoons, Frederick. 1991. *Food in China: A cultural and historical inquiry*. Boca Raton, Fla.
- Smith, Hedrick. 1991. *The new Russians*. New York.
- Spencer, Edward. 1899. *The flowing bowl*. Sixth edition. London.
- Tannahill, Reay. 1988. *Food in history*. New York.
- Toussaint-Samat, Maguelonne. 1987. *Histoire naturelle & morale de la nourriture*. Paris.
- Watney, John. 1976. *Mother's ruin: A history of gin*. London.
- Watson, Stanley A., and Paul Ramstead, eds. 1987. *Corn: Chemistry and technology*. St. Paul, Minn.
- Wigginton, Eliot, ed. 1972. *The Foxfire book: Hog dressing; log cabin building; mountain crafts and foods; planting by the signs; snake lore, hunting tales, faith healing; moonshining; and other affairs of plain living*. Garden City, N.Y.

III.6 🍷 Kava

Kava is both a plant and a drink made from that plant for ritual occasions. Kava usage is limited mostly to the Pacific basin, where it occurs widely from New Guinea in the west to the Marquesas in the east, and from Hawaii in the north to the southern Cook Islands. Some societies have ceased using it in recent times, whereas others ceased but began again after missionary prohibitions lessened and national independence brought kava to the fore as a mark of national identity (Brunton 1989).

A narcotic effect is commonly thought to be the main reason for kava's consumption, but elaborate rituals have developed with kava as their centerpiece, together with complex rules about who can drink the substance and when. Powerful cultural elements that persisted into the 1990s led to the commercialization of the root; kava is now drunk by overseas island communities in Auckland, Sydney, Honolulu, and Los Angeles. Indeed it seems that kava has evolved as a major force in the maintenance of the identities of Pacific islanders at home and abroad.

"Kava" is the term for the whole plant, which according to Western botanical terminology, is *Piper methysticum*, placing it among the pepper families (Lebot, Merlin, and Lindstrom 1992). But "kava" may

also refer to the beverage made from the roots or stem of the plant. Moreover, the term can mean a ritual in which the crushing of the root to make the beverage is a noteworthy activity.

Kava is a reconstructed Proto-Oceanic term within the large Austronesian language family (Biggs, Walsh, and Waqa 1972: 30). It, or a similar word, such as 'ava, is still widely used throughout the eastern Pacific to refer to the same root. In Pohnpei in Micronesia, the word is *sakau*, a close cognate. In Fiji, kava is referred to as *yaqona*, a term that probably reflects its bitter taste, *qona/kona*. Throughout Vanuatu, which has some 170 vernacular languages and dialects, the terms for kava are quite varied (see Lebot and Cabalion 1988: 54-67 for a list). In northern Vanuatu, kava is known as *maloku*, meaning quiet, subdued (Crowley 1990), while *nikawa*, a close cognate, appears only in Tanna in southern Vanuatu. The local term for kava in Papua New Guinea is also quite varied (see Lebot and Cabalion 1988; Brunton 1989).

Origin and Spread

The origin of the kava plant and the relationships among the many different cultivars found today have both been subjects of detailed study by botanists (Lebot, Aradhya, and Manshardt 1991). Using material collected from throughout the Pacific, they have traced genetic links among the various samples, concentrating particularly on diversity in the Melanesian area.

Vincent Lebot and P. Cabalion (1988) have produced an inventory of the names of local cultivars around the Pacific. From chromosome counts they distinguish the genus *Piper wichmannii* from that of *Piper methysticum*. *P. wichmannii*, the wild form, has been located only in Melanesia, that is, in parts of New Guinea, the Solomon islands, and northern Vanuatu. In contrast, *P. methysticum*, the cultivated form, has been found throughout the high islands of Polynesia and on the high islands of Pohnpei and Kosrae in eastern Micronesia, and in Melanesia. *P. wichmannii* has been found to have higher isozyme variability than *P. methysticum*, but the two overlap for zymotype 9, leading the authors to suggest that there is no taxonomic distinction between the two plants.

Lebot and Cabalion, along with Vincent Lebot et al. (1991: 181), suggest that what was being selected for was the kavalactones, which have a physiological effect when the product is drunk. They indicate that *P. methysticum* was "domesticated through vegetative propagation from a narrow genetic base in wild fertile progenitor . . . becoming sterile through mutations affecting fertility. . . . Because *P. methysticum* must be cultivated, this plant has resulted from human selection of somatic mutants. . . . That selection process has resulted in variability in both morphological characteristics and the kavalactones" (Lebot et al. 1991: 181). The root of *P. wichmannii*, which according to Vanu-

atu oral tradition was made into a drink, was too powerful in its effect and induced nausea (Weightman 1989: 236). This may be a reason that *P. methysticum* was developed instead.

Identification of these two closely related species leads botanists and a historical linguist to suggest that northern Vanuatu is the center of origin of kava cultivars, which have been domesticated probably in the last 2,500 to 3,000 years. From Vanuatu, varieties of kava were introduced to Papua New Guinea, eastern Micronesia, and Polynesia. This mainly west-east spread was followed by an east-west spread, from Tonga and Samoa to East Futuna, from which some Futunan speakers sailed westward to settle an island off south Vanuatu that they named West Futuna. From there it spread to the other Polynesian outliers and the islands in southern Vanuatu (Crowley 1990). According to Lebot et al. (1991: 184): "Kava is a relatively late introduction into Polynesia, since there is no variation in isozymes in that region." Further details of the genetic history of kava are given in Lebot, Merlin, and Lindstrom (1992).

Botanical Description

The kava plant is a perennial shrub that reaches one or two meters in height. It has many stems and bears a light foliage of heart-shaped leaves. The stems are notable for being distinctly segmented by a dark band, similar to bamboo.

The plant is dioecious; that is, it bears male and female flowers on separate plants. To propagate kava, a stem bud is taken from a *P. methysticum* male plant, and from this a root system and shoots develop. Vegetative reproduction, thus, allows a high degree of selectivity in which plants are selected for propagation. The importance of the plant as a ritual drink has led to considerable human intervention in its evolution with resulting diversification.

Kava is cultivated along with other root crops, such as taro and yams, in household gardens and on shifting agriculture plantations. Preferred locations vary: In Futuna in the 1960s, the plants were cultivated high on the mountainside because large ants attacked them nearer the coast (Gaillot 1962); today, however, kava is cultivated in the plantation area behind the houses of coastal areas (Pollock 1995a).

To harvest the root in Vanuatu, the plant is dug out after three to four years of growth when the root-stock and mass of fibrous roots may weigh 5 to 10 kilos. In Futuna and Tonga, the root with its stem still attached but leaves removed may be offered as a tribute at an important occasion, such as the induction of a chief. For example, in Tonga for the induction ceremony of the crown prince, some 300 stems and roots of kava were cultivated and donated by the families living on lands belonging to the Tongan royal family. In Samoa the roots may be dried before they are presented at a kava ceremony; each honored guest is pre-

sented with his (or her) stem of kava, so that some stems may change hands several times before they are finally pounded to make a kava drink.

Distribution

We surmise that many changes over time have affected the distribution of the plant, as well as the beverage and the ritual. Thus, in precontact times, all three are likely to have been different than those of today.

There is no reported evidence of the kava plant on New Caledonia, Easter Island, Belau, Yap, and parts of Papua New Guinea, nor on the atolls. In the case of the atolls, the soils are too poor and there is too much exposure to salt spray. In New Zealand it is likely that the climate has been too cold for the plant to grow.

Kava, as a drink, also shows no clear patterning. Records of its employment as a beverage come from those places that grew kava, which makes sense because the plant has to be vegetatively propagated and, consequently, was grown for a reason. But in places such as Tikopia, the plant continues to grow even though its usage has disappeared.

In those societies which do employ the plant ritually, a number of variations have developed over time and space. In some societies, these ritual occasions are still very formal, as in Futuna, Samoa, and Tonga, while in others a more secular consumption of kava has become popular, as in Pohnpei and Vanuatu. In Fiji, Tonga, Futuna, Pohnpei, and Vanuatu, both ritual and secular uses are practiced today.

According to missionary accounts, the nineteenth century saw the customary utilization of kava disappear in Tahiti, the Cook Islands, Hawaii, and Kosrae. This also happened in southern Vanuatu in the 1920s, but there kava drinking was revived after independence in 1979 (Brunton 1989). By contrast, on Wallis, widespread kava ceremonies declined after World War II and today are held only for very special occasions. Still another variation is found on the neighboring islands of Tikopia and Anuta, where in the 1930s kava was prepared as a drink. But the people did not drink it; rather, the ritual libation was poured to the gods (Firth 1970; Feinberg 1981).

Difference may also be found in the ways the root is prepared for drinking. In some societies it has been chewed (mainly by young women), while in others it is pounded or grated. Missionaries exerted a strong influence to discourage chewing kava because they considered it to be unhygienic and a way of spreading disease. In Vanuatu, young men have taken over the role of preparing the root, but in western Polynesia it still belongs to women, particularly on very important occasions.

Some societies process the root when it is green, whereas others prefer a beverage made from the dry root. In Fiji, at a highly formal welcoming ceremony, a whole fresh green plant is presented to the chief

guest. If the fresh green plant is unobtainable, then an appropriate amount of dried, or even powdered, root is presented (Ravuvu 1987: 25).

The patchy nature of kava use has intrigued anthropologists, botanists, and others for a century or more. Edwin Burrows, for example, in his comparison of cultural features of western Polynesian societies (which he saw as distinct from Melanesian and other Polynesian societies), noted that "the western Polynesian kava complex appears as a local elaboration on a widespread Oceanic base. The occurrence of the whole complex in parts of Fiji is probably due to diffusion from western Polynesia. . . . Distinct resemblances to western Polynesian kava customs elsewhere in Melanesia and Micronesia are also probably due to diffusion from western Polynesia" (Burrows 1939: 114-15). In western Polynesia he included Samoa, Tonga, 'Uvea/Wallis, and Futuna along with Fiji.

W. H. Rivers (1914) had earlier attempted to account for this scattered distribution by distinguishing "kava people" from "betel people," suggesting that exploitation of the two forms of the *Piper* plant were mutually exclusive. Betel is a combination of two plant substances, the *Areca catechu* nut and the *Piper betle* leaf, in which the nut is wrapped so that it can be chewed. The leaf contains the narcotic substances.

Ron Brunton (1989), in his society-by-society reexamination of Rivers's thesis, however, indicates that the distribution is not quite so mutually exclusive. His review of the linguistic and archeological evidence suggests to him that the Bismarck Archipelago is the "homeland" of kava, where it was part of an early social complex known as Lapita culture. From there it was traded along a coastal route to southern New Guinea and, later, to Fiji and Polynesia. He hypothesizes that kava moved with the eastward spread of Lapita culture some 3,000 years ago (Brunton 1989: 82). Such a spread was not unidirectional, and kava usage later diffused back westward from Samoa and Tonga. It, thus, reached southern Vanuatu, and Tanna in particular, he argues, several hundred years after it had become a part of the culture of northern Vanuatu. The most likely agents of transmission were the Polynesians of Aniwa and West Futuna, two islands just off the coast of southern Vanuatu (inhabited by Polynesian speakers), who brought kava customs developed in eastern Oceania back to Melanesia (Brunton 1989: 83).

Brunton also considers it likely that kava was once drunk by many peoples in Melanesia who subsequently abandoned it for unknown reasons. Similarly, he argues that betel, too, may have dropped out of use in some societies that may have subsequently exploited kava. Thus, he considers the belief that kava and betel are mutually exclusive to be largely mistaken. Both have been in and out of use over time in various parts of the Pacific (Brunton 1989: 85).

It can be concluded, therefore, that not every

island society has employed kava, that the plant does not occur everywhere, and that the ways in which it was employed also differ from place to place. Such fluctuations in occurrence of the plant and the customs associated with it are likely to have marked its long history. This may be because a particular society chose to drop such customs, or was forced to do so after cyclone damage rendered the plant unavailable, as happened in Futuna in 1986 (Pollock field notes 1987). Yet at a later time, the plant could have been reintroduced from a neighboring island along with new ways of employing it. Consequently, today we can see a range of kava usage from very ritual occasions to more secular ones, which may reflect differences in its social importance.

Ritual Uses

Kava rituals have become differentiated in both practice and ideology as societies have dispersed over time and developed their own distinctive cultural characteristics. The rituals took two broad forms: those in which the kava was prepared and drunk, usually in a kava circle, and those in which the whole root was presented to an honored person. But in either case, the root symbolizes the ties between ancestors and present-day peoples, represented by their chiefs and nobles; it, thus, symbolizes both past and present.

The most formal rituals at which kava is drunk begin with the appropriate people seated in an inner circle, with the community assembled around its outer edge. The processing of root to make the beverage is under the direction of the chief officiator for kava ceremonies. The ritual ends when the chief dignitary signals that the circle should break up - often when he himself has moved out of the circle.

Kava rituals take place outdoors as well as in specially designated houses. In Futuna, a house known as *tauas* is specially set aside for men's nightly kava sessions. For large formal kava ceremonies, at which many communities are present, an outdoor venue is obviously necessary. In Vanuatu, the kava-drinking grounds are outdoor sites, usually under banyan trees.

Processing

The processing of kava, which transforms the whole root into a mass of fibers and pulp, is also highly ritualized. It consists of three steps: pulverizing the root, adding water to the pulp, and serving the liquid. The main person making kava in western Polynesia is usually a young woman, who is aided by several young men.

Processing begins once everyone is seated and the necessary pieces of equipment are in place, including the wooden kava bowl, whole roots, water, the fiber strainer, and serving cups. The kava root, whether green or dried, is cut into small pieces and chewed,

pounded, or grated, according to local custom. In the past, chewing the root was the most widespread practice; this task was assigned to several young girls or young men, chosen for their perfect teeth. They washed out their mouths before commencing to chew until all the pulp was macerated, leaving a fibrous ball that each girl took from her mouth and placed in the kava bowl.

The mode of reducing the kava root to pulp has changed over time. As already mentioned missionaries discouraged chewing the root on the grounds that the practice spread disease. Alternative methods call for pounding the root, grinding it between two stones, or grating it. Upon the command of the chief of the kava ceremonies, water is added to the balls of pulp and fiber in the wooden kava bowl. The chief calls out the order as to when the kava should be mixed, how much water is to be added, and when the mixture should be strained. Adding the water is the most significant part of this ceremonial preparation. According to Futunans, it transforms one substance, the root, into another, the beverage. The root, thus, becomes the medium for communication with the gods, a means of both honoring and supplicating them.

The third stage begins once the mixture is ready to be served. Cups are carried to individual participants in the kava circle by young men designated for the task. The chief officiator calls the name of the person to be served. In Samoa these names are exclusively for use in the kava ceremony (known as kava cup names) and are not used in other situations (Williamson 1939).

Upon receipt of the cup, the recipient claps and pours out a small libation to the gods before drinking the remaining liquid in one gulp. In Samoa the first cup, designated for a particular god, used to be poured out by the chief officiator before others were served. The Tikopian practice of pouring out the carefully prepared liquid (Firth 1970) was thus in line with the general ritual procedures, but represented an alternative kind of development in that no living person drank the kava.

Clearly, such rites performed in front of the assembled community are very formalized. All movements and gestures have become stylized, especially in western Polynesia, where those eligible to perform the tasks undergo elaborate training.

Kava Drinking Circle

The seating order at each ceremonial event in western Polynesian societies is established by the kava officiator, who uses his knowledge of the relative status of the participants. The individual with the highest status sits in the center of the arc, with others seated to his left and right according to their relative status. In western Polynesian systems, a *matapule*, or talking chief, sits next to his chief. The circle may consist of between 15 and 40 persons.

The serving of the kava cups according to the order called by the chief officiator, thus, serves to reiterate the community hierarchy in a very visual and visible manner. The status of particular titles relative to one another and the incumbents holding those titles are displayed for community knowledge and affirmation.

Degrees of Formality

The most formal kava circles occur for the investiture of a new supreme chief. Such a ceremony was performed in Tonga in 1976 for the investiture of the crown prince of Tonga, an event that coincided with the Tonga Constitutional Centenary celebrations.

In Futuna and the neighboring island of Rotuma, a number of lesser rituals involving the kava plant were practiced, and some still are. A morning cup of kava was drunk together by chiefs of friendly hamlets to propitiate the spirits for the right outcome of the day's events (Rozier 1963). On 'Uvea/Wallis, by contrast, the range of kava occasions has been reduced to the very formal one for the installation of a new Lavelua (chief) and to some others associated with the Catholic Church's annual first communion celebration.

Although their island is nominally part of western Polynesia, Niueans have had a cultural ideology that differs from that of their neighbors. Edwin Loeb (1926) reported that only the priest was allowed to drink a potion to the gods, doubtless because the island's poor soils make it difficult to grow kava. Hence, its use was limited to priests.

Pohnpei, though in Micronesia, has shared much of the formality of western Polynesian societies, where kava (known as *sakau*) was used to support an elaborate chiefly system. Today, major events that draw the districts together are still marked by a kava ceremony, but much is also drunk informally (Peterson 1995).

In Vanuatu, only the island of Tanna has maintained the ritual use of kava. There, the emphasis is less on the hierarchy of the circle; the focal part of the ritual is when each man spits out the last mouthful of his cup by way of sending messages to the ancestors (Brunton 1989; Lindstrom 1987).

In other parts of Vanuatu, kava is drunk but with considerably less ceremony. The heaviest users have been the healers or controllers of magic, who are said to use the drink to improve their communication with the gods (Young 1995).

In those parts of Polynesia where kava rituals are no longer practiced, we can still glean a notion of their formality from the written accounts by Europeans. In Tahiti, kava was used only by chiefs, but not as part of any religious ritual (Oliver 1987). Similarly, the chiefs in the Marquesas were reported to follow the steps in ritual processing, but no elaborate ceremonies were practiced (Denning 1980). In the Cook Islands, reports indicate only that a bowl of kava was

made at an installation ceremony (Gill 1876), but we have no further detail.

In Hawaii, the practice of drinking *'awa* died out in the mid-nineteenth century. It was drunk mainly by the *ali'i* or chiefs, though it was not forbidden to commoners. As E. S. C. Handy summarized the practices: "Awa's ceremonial uses were simply the expression of the belief that gods like the same good things of life that men did" (1972: 191).

In earlier times, kava usage was also highly ritualized on Tikopia, a small island north of Vanuatu with a Polynesian culture, though geographically among Melanesian peoples. But it differed markedly from western Polynesian practices in one small but important detail. As Raymond Firth (1970: 203) noted, "it was poured out, not drunk." Otherwise its preparation and presentation followed the practices of other Polynesian societies. According to Firth: "the whole ideology of kava concentrated upon its religious significance in ritual. There were no social kava-drinking ceremonies, nor any consumption of kava as a beverage apart from [a very few] casual instances. In such a religious context the material properties of the kava were of less importance than its signalization function" (1970: 204).

Thus, ritual uses of kava across the Pacific vary considerably. Kava drinking almost everywhere has been predominantly the prerogative of chiefs and priests. Where other men are allowed to participate, it linked the more highly ranked senior men with the more lowly ranked. The main function of kava in these rituals is to communicate with the ancestral spirits, thus imbuing kava with an important mediating role. The whole ritual has a social rather than an individual character; it is a group experience.

Exchanges of the Root

Presentations of the kava root to a chief or an honored guest were ritually made to ask for a favor, or to atone for a wrong, or to ask a priest to propitiate the gods for a special service, such as stopping destructive winds (Rozier 1963). Moreover, at ceremonial functions where it was not appropriate to make a kava drink, whole roots might be presented to a visiting dignitary. Such a procedure still occurs in New Zealand within the Samoan community; a visiting dignitary from Samoa or elsewhere is presented with a root as a token of respect.

In Tanna, Vanuatu, kava was at the center of a network of exchanges that linked villages through "paths" or "roads" that connected the various kava-drinking grounds. Social relationships were developed and maintained along these roads, over which knowledge, information, and goodwill also passed (Brunton 1989; Lebot, Merlin, and Lindstrom 1992).

The kava root is also given in exchange for medicine or as a gift between friends. Such exchanges are

noteworthy because they demonstrate that the root itself can be more important than the narcotic properties it bears when made into a drink. In other words, it is not necessary for the root to be processed for it to have honorific meaning (Brunton 1989; Lebot, Merlin, and Lindstrom 1992).

Secular Usage

Kava has also been employed informally in Fiji, Futuna, Pohnpei, Tonga, and parts of Vanuatu where ritual preparations were minimized and there were fewer restrictions as to who could drink. Kava was drunk communally at designated meeting places, and kava sessions, lasting all night, were held several times a week. These sessions, although less ritually structured, nevertheless emphasized the symbolic nature of kava drinking and communication with the ancestors.

In Futuna, each village had (and has) its *tauasu* house for the men. In Tonga, the men may meet in the house of a prominent person. In Vanuatu, kava-drinking grounds were the nightly meeting places in villages, whereas kava clubs are the meeting place in towns. At these nightly sessions in Tonga or Futuna, the root is chewed or crushed by a young man or young woman, who is seated at the kava bowl in order to maintain the supply of the beverage by replenishing the cups of the drinkers and crushing more of the root as needed. The drinkers, all male, sit on mats in an oval around the edge of the house with the kava maker at one end. Such sessions can last until three or four in the morning, but some men quietly drift away as they become sleepy. There is a light buzz of conversation, and the occasional clap as a drinker receives his new cup (Pollock field notes, Tonga 1976).

Such occasions are mainly social, with conversation ranging around political issues and local affairs. In Tonga, Futuna, and Fiji, these sessions enable men to relax in the company of other men of the village and share information. Indeed, a young man is expected to attend if he wants to be viewed as interested in village affairs, and thus a candidate for a leadership position (Sepeliano 1974).

By the 1980s, kava drinking in Vanuatu was also a nightly occasion, though customs varied throughout the islands. The village gathering place, often under a banyan tree, gave way to a kava bar where women and young men also drink. The kava is drunk in one gulp, after which one is expected to spit loudly. Such bars have become very lucrative for their owners.

In Vanuatu certain varieties of the plant have been designated for everyday use, whereas other varieties are drunk only by persons of high rank, and still others are used as traditional medicine. Depending on the varieties available, together with the way they are prepared, kava can precipitate drunkenness,

but a bleary-eyed, staggering, and comatose sort, "never hilarious or pugnacious" (Gunn quoted in Weightman 1989: 239). In fact, with the custom of modern kava drinking in Vanuatu, there has arisen a whole range of slang phrases, such as *fowil antap*, "four wheel on top," to describe just one form of drunkenness.

Narcotic Properties

Kava has been labeled a narcotic containing certain pharmacologically active substances long recognized by Western chemists, pharmacists, and others. Certainly missionaries and other outsiders judged drug properties to be the reason for kava usage and took steps to ban its cultivation, processing, and use as a beverage. The main objection seems not to have been to drunkenness but to the soporific effects it produced, which prevented many men from doing a full day's work after a night of drinking kava. Recent ethnographic accounts from Vanuatu have stressed the druglike or narcotic properties of kava (Lindstrom 1987; Lebot and Cabalion 1988; Brunton 1989).

The main chemically active constituents identified by chemists are kawain and the kava lactones. However, despite many years of investigations, mainly in German laboratories (see Lebot and Cabalion 1988) but also in Sydney (Duffield and Jamieson 1991), the precise physiological action of these substances on the human neurological and chemical system is not fully understood.

Kawain is said to be an emotional and muscular relaxant that stabilizes the feelings and stimulates the ability to think and act. It has bactericidal properties and can be used as an antimycotic. It is also a diuretic. Fresh kava has a local anesthetic effect on the chewer's mouth. But its main effect is as a muscle relaxant (Lebot and Cabalion 1988: 35).

The chemical properties of kava and other local plants have long been of interest to visitors to the Pacific. Gilbert Cuzent, a naval pharmacist based in Tahiti from 1858 to 1860, claimed to be the first to identify (in 1858) a substance he called "Kavahine" as a result of experiments carried out on various parts of the kava plant. This claim led to a scientific argument with another French pharmacist, who had also published his analysis of the kava root in April 1857 in the newspaper *Le Messager de Tahiti* and, in 1858, in the *Revue Coloniale*. However, the French Academy of Sciences recognized Cuzent's claim (Cuzent 1860: 189-90).

The major physiological effects are quiescent and numbing, in contrast to the enervating effects of alcohol. The kava drinker may feel a slight numbness around the mouth, but the strongest effect is on the legs; anyone who sits drinking kava for a long period of time finds it hard to stand or walk. There is no loss of consciousness, though the kava drinker may fall asleep

after seven or eight cups and be hard to awaken. Some Vanuatu cultivars are more potent than others and so are more favored by drinkers for the quick effect they produce (Crowley 1990).

Kava is also said to lead to loss of appetite and to reduced libido, but such effects are reversible if the person stops drinking it for several weeks (Spencer 1941). Redness around the eyes is also a mark of a heavy drinker, as is a scaly skin.

In his studies of the plant's usage in the eastern Pacific, Edwin Lemert (1967) noted how kava produced a nonaggressive, anaphrodisiac, mildly tranquil and dreamy state. He suggested that it depressed bodily functions such as heart and respiration rates and temperature. He labeled kava drinking "a form of retreatist or avoidance behavior, related to onerous claims which Polynesian social organization periodically makes on individuals" (Lemert 1967: 337). Many authors have noted that the quiescent and soothing effects of kava place it in direct contrast to alcohol. For this reason, kava has been introduced as a counter to heavy alcohol drinking, whether in Vanuatu or among Australian Aborigines (D'Abbs 1995).

It is true that modern writers tend to think that those who drink enough kava become drunk, yet a number of other writers over the past century and a half have either not mentioned any drunkenness, or have said that kava does not lead to such a state and that drinkers only become sleepy and quiescent.

Kava and Alcohol

The relationship between kava and alcohol is one of contrast rather than one of similarity. As W. T. Wawn wrote in 1870: "Kava has a very different effect from alcohol. It is soothing, and a pint of strong kava, or even half that quantity for a beginner, will apparently have no more effect than to make a man desirous of being left alone and allowed to sit quietly and smoke his pipe. . . . Alcohol excites, kava soothes and then stupefies" (quoted in Weightman 1989: 237-8). Other authors have made similar observations. Thus, alcohol and kava satisfy different needs (Lemert 1967: 337).

One major difference between the two substances is that kava is a very social drink (it is almost unheard of for someone to pound a batch of kava for personal drinking), drunk in association with others in ritual settings or in the modern-day kava clubs, where a group of people share a bucketful (or one "brew") of kava.

Nonetheless, as a result of institutionalized drug-classification principles in the West, kava and alcohol have been placed alongside one another. Kava has been a banned substance in the United States since 1958, though there has been a campaign to lift that ban (Lebot and Cabalion 1988: 91). It is also considered a harmful substance in New Zealand, Australia, and Fiji.

Commercial Developments

Kava has been a cash crop for more than 100 years. Since the second half of the nineteenth century, several South Pacific countries have been exporting the dried root to Germany and France for pharmaceutical uses. In the 1880s, a trading house in Wallis sent 30,000 pounds in 18 months at a price of 30 to 35 cents a pound to meet a growing demand in Europe, where it was employed both as a diuretic and for its calming effects (Deschamps 1883).

Today, markets for kava in Germany and France still exist, and processing laboratories import the equivalent of some 200 tons of fresh root from Fiji and Tonga. A study of such demand showed a recognition of the therapeutic properties in kava for antiseptics, expectorants, diuretics, and urogenital stimulants. One product with the brand name Kaviase had been on the market for about 20 years, and in Germany an attempt was made to launch a kava-based soft drink (Weightman 1989: 241).

Certainly, kava has become an important alternative cash crop for several Pacific island societies, and for farmers in parts of Fiji and Vanuatu, it is now their most important cash earner. The crop is widely sold by the bundle of dried roots in the markets throughout Fiji. Purchasers are mainly urban Fijians who have no access to *yaqona* from their own lands. Similarly, in Vanuatu, kava already yields a higher net income per hectare to the farmer than cocoa and coffee (Lebot and Cabalion 1988: 92). It thus provides a form of economic buffer against downturns in tourism due to oil prices and political instability.

Conclusions

The evolution of kava usage over millennia has been marked by a waxing and waning of its popularity, as well as by changes in the plant and its mode of processing. During this time it has been eliminated from some islands by cyclones, salt water inundations, and even warfare when the plants were pulled out by the retreating forces (Gill 1876).

Yet vegetative reproductive properties have enabled varieties to be selected that could be processed into an ever more pleasant beverage, and the range of cultivars has increased markedly in island societies, such as those of Vanuatu where it is of great cultural importance.

The geographical range of kava, however, is more narrow today than it was a hundred years ago. Missionaries sought to eliminate it because it reduced productive work efforts, and other Westerners have included it under the negative rubric of a drug, although there is no medical evidence for long-term harmful effects.

The employment of kava in recent times has been marked by a considerable increase in its secular, as

opposed to its ritual, use. In Fiji, Futuna, Pohnpei, Tonga, and Samoa, the Catholic, Methodist, and Congregational churches have taken a less rigid stance against kava drinking, so the rituals have continued and broadened in scope to include welcoming ceremonies for Western visitors and such church activities as first communion. The kava parties now are open to men of all ages and status in the community, although women do not often participate.

Yet the overall development of kava throughout the Pacific during some 3,000 years has been one of ritual usage. Those societies that have maintained kava drinking to present times have done so because it is an important medium in their cultural system. It is steeped in a strong body of local ideology that places the root as central to a communication process, linking the ancestral spirits to the incumbent chiefs and priests, and, thus, the whole living community to those associated with its past.

Nancy J. Pollock

Bibliography

- Barrau, Jacques. 1965. L'Humide et le sec. *Journal of Polynesian Society* 74: 329-46.
- Biggs, B., D. Walsh, and J. Waqa. 1972. Proto-Polynesian reconstructions with English to Proto-Polynesian finder list. Working papers, Department of Anthropology, University of Auckland.
- Bott, Elizabeth. 1972a. Psychoanalysis and ceremony. In *The interpretation of ritual*, ed. J. S. La Fontaine, 205-37. London.
- 1972b. The significance of kava in Tongan myth and ritual. In *The interpretation of ritual*, ed. J. S. La Fontaine, 277-82. London.
- Brunton, Ron. 1989. *The abandoned narcotic*. Cambridge.
- Burrows, Edwin. 1939. *Western Polynesia*. Stockholm.
- Crowley, Terry. 1990. Who drank kava. Paper read at Conference on Austronesian Terminologies: Continuity and Change, Australian National University.
- Cuzent, Gilbert. [1860] 1983. *Archipel de Tabiti*, ed. J. Florence. Haere Po No, Tahiti.
- d'Abbs, Peter. 1995. The power of kava or the power of ideas? Kava use and kava policy in The Northern Territory, Australia. In *The Power of Kava*. Nancy J. Pollock ed. *Canberra Anthropology* 18: no. 1 and 2. Special Volume 166-83.
- Dening, Greg. 1980. *Islands and beaches*. Honolulu.
- Deschamps, L. 1883. Les Iles Wallis. In *Le tour du monde*. Paris.
- Douglas, Mary, ed. 1987. *Constructive drinking*. Cambridge.
- Duffield, J., and L. Jamieson. 1991. Development of tolerance to kava in mice. *Clinical and Experimental Pharmacology and Physiology* 18: 571-8.
- Feinberg, Richard. 1981. *Anuta: Social structure of a Polynesian island*. Laie, Hawaii.
- Feldman, Harry. 1980. Informal kava drinking in Tonga. *Journal of the Polynesian Society* 89: 101-3.
- Firth, Raymond. 1970. *Rank and religion in Tikopia*. London.

- Gaillot, M. 1962. La rite du kava Futunien. *Études Mélanésiennes* 14-17: 95-105.
- Gill, W. W. 1876. *Life in the southern isles*. London.
- Handy, E. S. C. 1972. *Native planters in old Hawaii*. Bishop Museum Bulletin No. 233. Honolulu.
- Jamieson, Dana, and P. Duffield. 1991. Production of tolerance to the aqueous extract and lipid soluble extract of kava in mice. Abstract.
- Kirch, Patrick, and D. Yen. 1982. Tikopia: The prehistory and ecology of a Polynesian outlier. *Bishop Museum Bulletin* #238. Honolulu.
- Lebot, Vincent, Mallikarjuna Aradhya, and R. Manshardt. 1991. Geographical survey of genetic variation in kava. *Pacific Science* 45: 169-85.
- Lebot, Vincent, and P. Cabalion. 1988. *Kavas of Vanuatu*. South Pacific Commission Technical Paper No. 195. Noumea, New Caledonia.
- Lebot, V., M. Merlin, and L. Lindstrom. 1992. *Kava: The Pacific drug*. New Haven, Conn.
- Lemert, Edwin M. 1967. Secular use of kava in Tonga. *Quarterly Journal of Studies on Alcohol* 28: 328-41.
- Lindstrom, Lamont. 1987. Drunkenness and gender on Tanna, Vanuatu. In *Drugs in Western Pacific Societies: relations of substance*, ed. L. Lindstrom. Lanham, Md.
- Loeb, Edwin. 1926. History and traditions of Niue. *Bishop Museum Bulletin* #32. Honolulu.
- Marshall, Mac. 1979. *Beliefs, behaviors, and alcoholic beverages*. Ann Arbor, Mich.
- Newell, W. H. 1947. The kava ceremony in Tonga. *Journal of the Polynesian Society* 56: 364-417.
- Oliver, D. 1987. *The Pacific Islands*. Honolulu.
- Petersen, Glenn. 1995. The complexity of power, the subtlety of kava: Pohnpei's *sakau*. In *The Power of Kava*, Nancy J. Pollock ed. *Canberra Anthropology* 18: special volume nos. 1 and 2, 34-60.
- Pollock, Nancy J. 1995a. Introduction: The power of kava. In *The Power of Kava*. Nancy J. Pollock, ed. *Canberra Anthropology* 18: special volume nos. 1 and 2, 1-19.
- 1995b. The power of kava in Futuna and 'Uvea/Wallis. In *The Power of Kava*. Nancy J. Pollock, ed. *Canberra Anthropology* 18: Special volume nos. 1 and 2, 136-65.
- Quain, Buell. 1948. *Fijian village*. Chicago.
- Ravuvu, Asesela. 1987. *The Fijian ethos*. Suva.
- Rivers, W. H. 1914. *The history of Melanesian society*. 2 vols. Cambridge.
- Rossille, Richard. 1986. *Le kava à Wallis et Futuna*. Collection Iles et Archipels No. 6. Bordeaux, France.
- Rozier, Claude, ed. 1963. *Écrits de S. Pierre Channel*. Publication de la Société des Oceanistes no. 9. Paris.
- Sepeliano, T. 1974. Enquête sur la reproduction du mode de production capitaliste dans les îles Wallis et Futuna. Mémoire pour la maîtrise de sociologie, Université de Lyon II.
- Spencer, D. 1941. *Disease, religion, and society in the Fiji Islands*. Seattle, Wash.
- Weightman, Barry. 1989. *Agriculture in Vanuatu*. Cheam, Surrey.
- Williams, Thomas. [1858] 1982. *Fiji and the Fijians*. Suva.
- Williamson, R. W. [1939] 1975. *Essays in Polynesian ethnology*, ed. R. Piddington, 51-112. New York.
- Young, Michael. 1995. Kava and Christianity in Central Vanuatu: with an appendix on the ethnography of kava drinking in Nikaura, epi. In *The power of kava*, Nancy J. Pollock, ed. *Canberra Anthropology* 18: No. 1 and 2 special volume, 61-96.

III.7 Khat

Khat (*Catha edulis* Forsk., Celastraceae) is a flowering evergreen tree that grows in parts of eastern Africa and the southwestern highlands of Arabia. Its young leaves and tender stem tips, also called khat, are stimulating and produce a mild euphoria when chewed and their juices ingested. Khat has various psychostimulant alkaloids, of which the main one, cathinone, is amphetamine related and highly unstable. Khat leaves and shoots contain various alkaloids, of which cathinone and cathine are the primary and secondary active ingredients in stimulating the central nervous system. As with other psychoactive plants, the effects of the active constituents differ from one specimen to another according to the cultivar's size, age, health, and the site conditions of growth, such as exposure, soil, soil moisture, and drainage.

Most of the world's khat is grown in Yemen, Ethiopia, and Kenya, where its use is long established. Khat is a major import of Somalia and Djibouti, and demand in both countries far exceeds domestic production. During the past half-century, the number of users has increased substantially, and emigrants from the khat-chewing countries have introduced (on a small scale) the use of the drug to some countries in Europe and North America.

International trade in khat is a multimillion-dollar business in the Horn of Africa and southwestern Arabia. Daily consumption of the drug is estimated to be about 5 million portions of khat (Kalix 1994: 69), or about 500,000 kilograms of the leaf material chewed. The various estimates of the total number of regular users range from 3 to 6 million. Methcathinone, a synthetic, white, chunky powder, known in the United States by the street name "cat," is a substance very dissimilar from khat.

Orthography

"Khat" (pronounced "cot") is an inexact transliteration of the Arabic name of the plant, and is also spelled "qat," "kat," "gat," and "ghat." Philologists consider that the spellings "qāt" or "kāt" (with diacritical marks) most closely approximate the transliteration of the Arabic in Roman characters.

The Khat Plant

Botanical Description

There are two known species of the genus *Catha*: *Catha edulis* and *Catha spinosa*. Both were first described by Peter Forsskal in 1775. Only *C. edulis* is cultivated, and in eastern Africa and the southwestern Arabian peninsula it grows at higher, wetter elevations than *C. spinosa*, which occurs wild in northern Yemen.

Khat is a straight, slender tree with a thin bark that varies in color from light gray to dark brown. The

trees bear persistent, leathery leaves that are elliptical with finely toothed edges. The leaves are approximately 50 to 100 millimeters (mm) long and 20 to 50 mm broad (Revri 1983: 37). Young leaves are shiny and vary in color from pale green to red, becoming green or yellowish green at full growth. The flowers are small (diameter 4 to 5 mm) and either white or greenish in color. Freshly cut twigs that carry tender new leaves and buds near the end of the branch are most highly appreciated for chewing.

The height of mature trees under cultivation ranges from about 2 to 10 meters (Brooke 1960: 52; Kennedy 1987: 177). The leaves grow rapidly if the plant is well watered, and if heavy harvesting continues throughout the year, the trees retain the form of bushes (Tutwiler and Caprisco 1981: 53). Under very favorable conditions in eastern and eastcentral Africa, wild *C. edulis* grows to a height of about 25 meters (Greenway 1947: 98).

Khat requires a yearly minimum of approximately 400 mm of water from rainfall or irrigation and an average temperature not less than 17 degrees Celsius. Growth is retarded by successive night frosts at high elevations and by rainfall in excess of 800 mm if the soil is not freely drained (Revri 1983: 18-19). Healthy trees may yield for at least 50 years.

C. edulis has been cultivated in greenhouses at various botanical gardens and has been grown in the open for scientific purposes in Egypt, Sri Lanka, Bombay, Algeria, Portugal, southern France, Florida, and southern California.

Cultivation

Khat is reproduced vegetatively from shoots (suckers or sprouts) cut from the basal root or the trunk of the tree. These provide more rapid growth than cuttings from branches. Three or four shoots, 25 to 50 centimeters (cm) long, are planted together in rows of shallow holes at 1 to 2 meter intervals (Revri 1983: 62). Harvesting of leaves usually begins 3 to 4 years after planting. (Brooke 1960: 53; Morghem and Rufat 1983: 215-16).

Kinds of Khat

The total number of the kinds of khat consumed throughout the world, if known, would well exceed 100. There are at least 7 kinds marketed at Dire Dawa, Ethiopia, and more than 40 kinds are recognized in Yemen on the basis of geographical origin (Krikorian 1984: 160). Consumer differentiation among the various kinds is based on the potency of psychostimulation, flavor, and the tenderness of the leaves. These characteristics differ according to growing conditions, husbandry, and farm practice, such as irrigation or lack of it, soil types, and other ecological variables. According to Bob Hill (1965: 16): "It is said that the palates of some chewers are so sensitive that they can distinguish not only in what area the products originate, but the individual field as well."

The health of the plant is also a marketing factor. *C. edulis* is subject to attack by a wide range of insect pests, and a kind of khat called *kuda*, when damaged by a leafhopper (*Empoasca* spp.), is considered to be the best quality in the Alamaya district of Harer. A toxin contained in the saliva that the insect injects into leaf tissue during feeding produces a milky taste that is preferred by connoisseurs of this expensive kind of khat (Hill 1965: 21).

Khat is sold at the retail level in leafy bundles (*rubtab* in Arabic), each bundle wrapped in banana leaves or plastic to preserve freshness. A rubtah contains the minimum quantity of leaves and stem tips that most consumers chew in one day (Weir 1985: 89; Kennedy 1987: 80).

Historical Overview

Uncertain Origin of Domesticated Khat

Neither khat's place of origin nor the manner in which it was diffused has been determined, although most experts accept the view that *C. edulis* was probably first domesticated in the Ethiopian highlands. Based upon his field studies in Ethiopia, Eritrea, and Somalia in 1927, botanist and plant geneticist N. I. Vavilov (1951: 37-8) designated Ethiopia as the center of origin of cultivated khat. According to anthropologist George P. Murdock, khat and coffee are two of some eleven important agricultural plants originally domesticated by the Agau, Cushitic speakers of the central highlands of Ethiopia and "one of the most culturally creative people on the entire continent" (1959: 182-3).

Revri was among the minority of khat specialists who challenged the belief of an Ethiopian genesis of the plant. In his view, Yemen is probably the primary center of origin of khat and Ethiopia the secondary center. On the basis of cytogenetic evidence, Revri concluded that *C. edulis* appears to have evolved from *C. spinosa*, found wild in the Serat Mountains of Yemen. He suggested that *C. edulis*, known to the Arabs as a medicinal plant, may have been taken to Ethiopia in the sixth century A.D. and that it was returned to Yemen as a social stimulant in the fourteenth century (Revri 1983: 4).

Historical Highlights

According to Armin Schopen (1978: 45) and others, the earliest reference to khat is found in *Kitab al-Saidana fi al-Tibb*, a work on pharmacy and materia medica written in A.D. 1065 by the scholar Abu al-Biruni (Abu r-Raihan Muhammad al-Biruni) in collaboration with a physician. The work described khat as "a commodity from Turkestan. It is sour to taste and slenderly made in the manner of *batan-alu*. But qat is reddish with a slight blackish tinge. It is believed that batan alu is red, coolant, relieves biliousness, and is a refrigerant for the stomach and the liver."

Batan-alu is presumably an extract of khat pre-

pared like preserves of fruits and vegetables (Schopen 1978: 45; Krikorian 1984: 136). The fact that al-Biruni, who lived in Ghazni, Afghanistan, identified "Turkestan" (the name loosely applied to the large area of central Asia between Mongolia on the east and the Caspian Sea on the west) as the origin of *batan-alu* is interesting because this and other references to either wild or cultivated *C. edulis* in the region are uncorroborated. It is plausible that khat used in the preparation of *batan alu* were dried leaves grown in Ethiopia and brought to central Asia as one of the myriad trade articles carried along the great caravan routes of the Old World.

The earliest work in which khat was identified as a plant is a book of medicinal remedies written in A.D. 1222 by Nagib ad-Din as-Samarkandi. The author, who resided in the ancient city of Samarkand, a major center of trade in Turkestan during the Middle Ages, recommended khat for its healing properties.

The earliest reliable reference to the general practice of khat consumption is in a chronicle of the wars between the Muslim and Christian states of Ethiopia during the early fourteenth century. The chronicler, a Christian, mentioned that khat was popular and widely consumed among the Muslim population (but shunned by Christians). He also wrote that a Muslim sovereign, Sabr ad-Din, boasted of what he would do when he conquered the Christian realm of King Amda Syon: "I will take up my residence at Mar'adi, the capital of his kingdom, and I will plant chat there, because the Muslims love this plant" (Trimingham 1952: 228).

According to John G. Kennedy (1987: 60-78), the popular use of khat in Yemen probably began in the southern part of the western highlands near Ta'izz during the fourteenth or fifteenth century. Customary use of the plant spread slowly northward, and by the end of the eighteenth century, khat was regularly sold in San'a. During the nineteenth century, use of the drug was widespread in the country, and Yemenis began to export khat overland to Aden (the annual trade grew from 1,000 camel loads in 1887 to 2,000 in 1900). Greenway (1947: 99) mentioned that the use of the leaves of *C. edulis* was well known to Ethiopians, Somalis, and Arabs at the time of World War I and that the knowledge of such use was spread to other African tribes through their encounters with the khat-chewing custom during war service in Ethiopia and Kenya.

The earliest concise account of khat in European literature was that of the Swedish physician and botanist Peter Forsskal (1736-63), who was a member of a Danish expedition led by the German geographer Karsten Niebuhr that visited Yemen in 1763. Khat, given the name *Catha edulis*, was among the plants collected in Yemen. Niebuhr, the only survivor of the five members of the expedition, published Forsskal's botanical papers in 1775, and in memory of his friend extended the name *Catha edulis* to "Catha edulis Forsskal" (Revri 1983: 3).

Religious Role of Khat

In most of the legends and early historical accounts of khat, a common theme has been its capacity to enhance wakefulness and, therefore, the ability of the user to carry out religious observance and worship (Trimingham 1952: 228). A frequently heard comment in praise of khat by Muslim users is that it enables them to pray without becoming drowsy even throughout the nights of Ramadan (Brooke 1960: 53). In Ethiopia and in Yemen, stories are told of divine guidance in the discovery of khat and the high regard in which the drug was held by Muslim saints.

A well-known legend tells of Sheikh Ibrahim Abu Zarbay (Zerbin), one of 44 Muslim saints who came from the Hadhramaut (eastern Yemen) to Ethiopia in about A.D. 1430 on a proselytizing mission. He traveled to Harer, converted many to Islam, and is said to have introduced khat to Yemen upon his return (Burton 1910: 66-7).

Khat has long been known in predominantly Christian northern Ethiopia, although it was consumed there almost exclusively by the Muslim minority. Frederick Simoons (1960: 115-16) found that nearly all adult Muslims in the town of Gondar chewed khat, but that among the Wayto of Lake Tana only Muslim holy men did so. In Begemder and Semyen, Christians who are supposedly possessed by Muslim evil spirits use it to appease those spirits and to encourage them to leave (Simoons 1960: 115-16).

In the course of some 400 years (thirteenth to seventeenth centuries) of protracted struggle for territory between Christian and Muslim states in Ethiopia, the Christians came to identify the eating of khat as a distinguishing characteristic of the Muslims and disdained its use. However, in Yemen where Moslems and Jews have regularly used the drug, khat did not become an object of religious identification.

Yemeni Jews have used khat since at least the seventeenth century. Sholem bin Joseph al-Shibezi (1619-86) is the author of a poetic play in which a dialogue between coffee and khat is presented (Krikorian 1984: 151-2, citing J. Kafih 1963: 224-5). The play is still performed in Arabic by Yemeni Jews in Israel (Weir 1985: 75).

Within Islam, khat is at the center of controversy. Richard Burton observed during his visit to Harer, Ethiopia, in 1855 that khat produced "a manner of dreamy enjoyment and the Ulema [authorities in Muslim law and religion] as in Arabia, held the drug to be 'Akl el Salikim,' or the Food of the Pious." Burton also wrote that the literati thought khat had "singular properties of enlivening the imagination, clearing the ideas, cheering the heart, diminishing sleep, and taking the place of food" (1910: 232).

Such an expression of esteem for khat, however, represents a minority view in the Islamic world, and for centuries, its consumption has been the subject of debate by doctors of Islamic law. Khat is not mentioned in the Koran, and the persistent question is:

Does the use of khat contravene the Koran's general injunction against the use of intoxicants? Although not of one mind on the question, Muslim religious leaders in most of the Islamic countries have taken the position that it does.

The following resolution was adopted by the World Islamic Conference for the Campaign against Alcohol and Drugs that met at Medina, Saudi Arabia, in May, 1983:

After reviewing reports submitted to the Conference on the health, psychological, ethical, behavioral, social and economic damages resulting from khat, the Conference judges khat to be a drug prohibited by religion and accordingly the Conference recommends to Islamic states to apply punishment of the basis of Islamic Shari'ah [canon law] against any person who plants this tree and markets or consumes khat (Al-Hammad 1983: 228).

Earlier, in 1971, the government of Saudi Arabia had banned the importation or use of khat in the kingdom and prescribed severe penalties for violations.

In the Muslim countries where khat is legal, there is an ambivalent attitude toward its use, and opposition to the drug is not necessarily based on religion. Many intellectuals in these countries deprecate the use of khat on the grounds that it is a deterrent to economic and social progress (Weir 1985: 66).

Although, traditionally, khat has appealed more to Muslims than non-Muslims, its use in the Islamic world is and has been minute. As A. D. Krikorian (1984: 163) has pointed out, the Turks in Ottoman-occupied Yemen never adopted the practice, nor did the 60,000 Egyptian soldiers stationed in Yemen from 1962 to 1967. In fact, less than half of one percent of the world's Muslims use the drug.

Geographical Perspective of Khat

Asia

Yemen. Khat is the paramount crop of Yemen, where it is widely consumed and its use is legal. According to The Economist Intelligence Unit Limited (1994-5: 47-8): "It is hard to overestimate the social and economic importance of the stimulant shrub, *qat*. The majority of Yemenis, women less than men, chew the drug most days from early afternoon to evening. The habit is less prevalent in the southern governates and is rare in the eastern regions."

Data on khat are largely unrecorded in Yemen's official statistics, and quantitative assessments are estimates. Some aspects, however, are palpable. According to Tutwiler and Caprisco (1981: 52): "Compared to other perishable crops khat has the highest market value, the lowest water requirements, and demands the least output of heavy labor." Beyond all question, khat is the preeminent cash crop of the country.

In fact, a 1992 report of the *Yemen Times* suggested that the value added by the khat sector of Yemen's economy is equivalent to approximately a quarter of the recorded gross national product and about twice the value resulting from cultivation of all other crops (The Economist Intelligence Unit Limited 1994-95: 56). The khat industry of Yemen involves landowners and growers, pickers, packers, transporters, wholesalers, and retailers - some 500,000 people, equivalent to about 20 percent of the working population (The Economist Intelligence Unit Limited 1994-95: 56). A substantial part of government revenue is derived from taxes on khat, although there is little doubt that a large part of the crop escapes levy (The Economist Intelligence Unit Limited 1994-95: 56).

Until the 1970s, khat was too costly to be used frequently by most of the population. But concurrent with the Gulf oil boom that began in 1973-4, an unprecedented number of workers (more than a third of the potential male labor force in North Yemen) migrated for temporary employment to Saudi Arabia and other petroleum-producing countries in the Gulf (Varisco 1986: 2). Remittances sent home by the migrants (annually about 600 to 800 million dollars in the 1980s) created new wealth and a vastly increased demand for khat at home (Varisco 1986).

The early 1980s were exceptionally profitable years for khat farmers. A survey by Shelagh Weir (1983: 67) of a khat-growing community of 4,000 people in Razih Province, North Yemen, near the Saudi Arabian border, found that the market value for land yielding two harvests of khat per year was the equivalent of 90,000 U.S. dollars per hectare. Terraced land, on which two to three harvests of khat per year are common, sold for the equivalent of 200,000 to 600,000 dollars per hectare (Weir 1983: 67). Most holdings range between 0.15 and 0.5 hectares. In 1980, in the district of Rada, a major area of khat production, the estimated net profit of a khat farmer in the second year after the crop matured was the equivalent of about 37,000 dollars per hectare (Varisco 1986: 4).

Critics of the institutionalized use of khat have argued that the transcendent position of *C. edulis* in agriculture has depressed the production of other crops, including the staple grains, and that the large expenditure for khat drains family income that would better be spent on food. Indeed, since the 1970s, importation of foodstuffs has greatly increased in Yemen, and highly processed imported foods have brought changes to the traditional diet (Nyrop 1985: 108). Cereal imports increased 133 percent between 1987 and 1990 and accounted for 16 percent of total imports, compared with 11 percent in 1987 (The Economist Intelligence Unit Limited 1992: 39).

The economy suffered a severe blow, however, when in late 1990, the government of Saudi Arabia expelled all Yemeni migrant workers, terminating Yemen's main source of foreign exchange. The sudden return of some 850,000 Yemenis to their homeland,

and the additional loss of about 1 billion dollars annually in foreign aid from the Arab oil-producing states, had drastic effects on the economy. In mid-1994, unemployment was about 25 percent, and the rate of inflation was about 100 percent annually (The Economist Intelligence Unit Limited 1994-95: 48-9).

Although export of khat from Yemen had ceased in 1974, there are unofficial reports that the depressed economy has prompted efforts in Yemen to increase its export of khat by air to markets in Africa and Europe.

North Yemen (the former Yemen Arab Republic). According to Kennedy, who directed an extensive program of team research on the sociomedical aspects of khat in Yemen during the mid-1970s, "No other society is so influenced by the use of a drug as is North Yemen by the use of qat" (1987: 78). He added that from 80 to 90 percent of adult men and 30 to 60 percent of adult women chew khat more than once each week.

About 70 percent of the country's khat is produced in three areas of the western highlands: the district of San'a; the province of Ibb, Yemen's overall leading agricultural region; and Jabel Sabr (elevation 3,005 meters), the terraced mountain that overlooks the town of Ta'izz and is the oldest and most famous khat-growing area in Arabia (Revri 1983: 16). The greater part of Yemen's khat is consumed in the western interior mountains and plateaus where the majority of the population is located, but freshly harvested leaves are trucked daily from the high elevations, where it is grown, to settlements in the hot, dry coastal plain bordering the Red Sea.

Chewing during the long afternoons in gatherings of friends and acquaintances is a ritual in Yemeni culture that provides a focal point for social contact, informal business dealings, discussion of current events, mediation of disputes, and the free exchange of ideas in a friendly, relaxed milieu. In general, the ubiquitous afternoon "khat party" is a social ritual that underscores the institutionalized role of khat in Yemeni culture and accounts for the major part of the consumption of the drug in the country. Most such parties take place in private residences. They are "open house" affairs, and anyone may participate.

The houses of wealthy urban families have a special room for khat parties called *al Mafraj*, "a place for joyful gatherings" (Weir 1985: 110-11). As a paradigm, it is located with a pleasing overlook on the highest floor of the house, and a large window, extending almost to floor level, provides a good view for the seated guests. Yet, it is the custom that the door and the windows of the *mafraj* be closed during a khat party. During even ordinary sessions, the room is usually crowded with participants who exult in the oppressive heat and humidity and the smoke from cigarettes and the waterpipe. Users believe that the use of tobacco is essential to the enjoyment of khat

chewing (Kennedy 1987: 86). Because khat inhibits urination and defecation by the constriction of muscular vessels, chewers have the ability to remain seated throughout the session for the customary 4 or 5 hours (Kennedy 1987: 115).

The societies of the khat-chewing countries are segregated by sex, and although men are the main users of the drug, women constitute a significant minority of users. The widespread popularity of khat with women is a fairly recent phenomenon. During the 1970s and 1980s, a period of unprecedented prosperity in Yemen, the domestic chores of both urban and rural women were greatly reduced by technological innovations, such as powered flour mills, piped water supply, and the availability of motor transport. The time available to women for social gatherings was substantially increased, and at a growing proportion of these gatherings, khat was consumed (Weir 1985: 90-1).

South Yemen (the former People's Democratic Republic of Yemen). The strong market for khat in South Yemen is only to a small extent supplied by local production, and (despite some interruptions) for more than a century and a half, the farmed land near Ibb in North Yemen has been the main source of South Yemen's khat. From this major center of production, the drug is transported to markets in the port of Aden (population about 420,000), Yemen's commercial capital.

Until about the time of World War II, khat was transported from Ibb to Lahej by camel caravans, which were subject to pillage en route. Armed raiders pulled the camel loads to pieces in order to "extract a few choice bundles [of khat], with the result that the loads arrived in Aden dried up and unfit for human consumption" (Ingrams 1966: 106).

The government adopted an anti-khat policy, and in 1977, the use and sale of the leaf were limited by law to one day per week (Friday). This restriction, however, ended in 1990 when the two Yemens united.

Saudi Arabia. The small province of Jizan in the southwestern corner of Saudi Arabia is the only part of that country in which consumption of khat is of consequence. The mountain tribespeople in Jizan, settled in rugged highland terrain and in relative isolation, have historically maintained a large measure of autonomy in their internal affairs. The growing and usage of *C. edulis* is centuries old in the uplands of the province, but at present, because of the government's determination to abolish the plant and its use, cultivation of khat exists chiefly, if at all, covertly in the Fayfa area of Jizan. Schopen (1978: 65) identified Jebel Fayfa as the northern limit of khat cultivation in Arabia.

Israel. In the nineteenth century, khat growing was introduced by Yemenite Jews to the region of

Ottoman-ruled Palestine that later became the modern state of Israel (Erich Isaac 1995, personal communication). Large-scale migration of Jews from Yemen began in 1882. By 1948, when Israel was established, about a third of the Jewish population of Yemen had emigrated. The exodus of almost all Jews remaining in Yemen (about 50,000) was accomplished by air transport to Israel, beginning with "Operation on Eagles' Wings" in 1949 (*Jerusalem Post* September 19-25, International Edition 1982: 20).

It was a common practice of Yemeni migrant families to bring shoots of "gat" (*C. edulis*) with them to be planted in home gardens of their new settlements, and in Israel, the cultivation and consumption of khat remain almost exclusively the practices of Yemenite Jews. Immigrant Ethiopian Jews are reported to chew the leaf, as do some young Ashkenazis, who are attracted by the exotic nature of the drug (High on gat 1996: 16). Most khat is consumed by its growers, but the fresh leaves can be found for sale (legally) in some of the open-air "oriental markets" (Arnon Soffer, 1995, personal communication; D. Hemo, 1995, personal communication). A khat-based frozen concentrate, "Pisgat," is made and sold in Israel as a health food. Its maker is reported to claim that the psychotropic effect of two tablespoons of Pisgat (mixed with water or soya milk, or added to ice cream) is equivalent to that achieved by several hours of chewing fresh khat (High on gat 1996: 17). In 1952, the ministry of health reported to the Knesset that although *C. edulis* was a stimulant, there was no reason to ban the leaf; since then, this official view has prevailed.

Afghanistan and Turkestan. In scientific literature, there are a few brief uncorroborated references to the occurrence of the khat plant in Afghanistan and "Turkestan," that is, southcentral Asia. One reference, the most credible, is a two-paragraph communication titled "Catha Edulis," by F.J. Owen, published early this century in a British chemical journal. Owen, an analytical chemist employed by the government of Afghanistan, reported his observations of the use of khat as a beverage by Afghans in Kabul:

The plant is found in the south of Turkestan and certain parts of Afghanistan to the east of Kabul. . . . After inspecting a specimen brought to me for analysis I recognized it as the *Catha edulis* plant. The natives say that men by its aid can do long marches at night without feeling the least fatigue, also that among the wrestling fraternity here it is used on a large scale, as it greatly increases the muscular powers of the men. . . . It is drunk by many Afghans as a substitute for tea (1910: 1091).

Although more than 80 years have passed since the publication of Owen's communication, however, the existence of khat in Afghanistan (or in any other country in southcentral Asia) is still unconfirmed.

Africa: General Distribution

C. edulis occurs wild and as a cultivated plant in eastern and eastcentral Africa. Wild stands grow sporadically in highlands from the northeastern Horn of Africa to the Sneeuwege Range in the far south of the continent. The plant is reported to be found in Eritrea, Ethiopia, Somalia, Kenya, Uganda, Tanzania, Rwanda, Burundi, Zaire, Zambia, Malawi, Mozambique, Zimbabwe, and the Republic of South Africa.

Ethiopia. Khat is grown in at least 9 (of the 14) administrative areas (provinces) of Ethiopia (Revri 1983: 5). The plant is a prominently cultivated crop in the plow- and grain-farming complex of Harerge and eastern Arsi provinces and is frequently part of the planting complex of the Kefa and the Galla (Oromo) ethnic groups in the Kefa Province where khat ranks second to coffee as a cash crop (Westphal 1975).

The main area of khat cultivation is the Harer Plateau - in the eastern section of Ethiopia's Central Highlands - with commercial centers of production in the administrative subdistricts of Harer, Webera, Garamulata, Chercher, and Dire Dawa (Assefa 1983: 73). All of these are within a radius of 100 kilometers (km) from the city of Harer, where according to local tradition, khat was first domesticated. In any case, its cultivation and use at Harer is older and economically more significant than in other areas in Ethiopia.

Small farm villages are scattered throughout these highlands of Harer, and at elevations between 1,500 and 2,100 meters, climate and soils are favorable for the production of a wide variety of field, garden, and tree crops. Rotation and terracing are practiced, as is irrigation where possible. Durra grain sorghum (*Sorghum vulgare*) is the predominant food crop, and khat is the most valuable cash crop (Brooke 1958: 192) - replacing coffee, which held that position until the middle of the twentieth century. Coffee fell from favor with Hareri farmers after World War II, chiefly because of low prices fixed by the government.

In 1949, Ethiopian Airlines introduced commercial air transportation of freshly cut khat from Ethiopia to Djibouti (French Somaliland) and to the British Crown Colony of Aden (now South Yemen). Transport of the perishable leaves by air was a commercial success from the start. However, the military junta that deposed the Ethiopian emperor, Haile Selassie, and held power during the years 1975 to 1991, sought to suppress its cultivation. Strong measures were employed to achieve that goal (Rushby 1995: 17), but since the end of civil war and the inauguration of a new coalition government in 1991, khat production has resumed without interference.

The export of khat officially earned 16 million U.S. dollars in 1993-4, establishing the drug as Ethiopia's fifth-largest source of revenue (Rushby 1995: 17). Illicit trade in the leaf (smuggling and by other means

avoiding excise) probably generates at least as much revenue as the official figures.

Khat is airfreighted twice weekly from Dire Dawa to London and Frankfurt, where it is sold to Ethiopian, Somali, and Yemeni expatriate communities. A small bundle of the leaves, sufficient for two or three hours of chewing, sells in Germany and the United Kingdom for the equivalent of between 6 and 7.50 U.S. dollars (Rusby 1995: 16–17). Ethiopian exports of khat to Europe compete with those from Kenya and Yemen.

Djibouti. The Republic of Djibouti (population about 400,000) is a major consumer of khat. Almost all of the drug is airfreighted from Dire Dawa, Ethiopia, where international cargo flights with freshly cut khat were inaugurated in 1949. The Economist Intelligence Unit Limited (1989: 56) has estimated that as much as 12 tons of khat, valued at about 40 million U.S. dollars, enter Djibouti daily.

Somalia. Indigenous production of *C. edulis* in Somalia is small, and most of the khat consumed in this country is imported from Kenya and Ethiopia. Prior to the creation of the Somali Republic by the merger in 1960 of the (British) Somaliland Protectorate and Italian Somaliland, khat, on a small scale, was mainly used in the city of Hargeysa, the administrative center of the British dependency. After World War II, khat use in the protectorate increased considerably in spite of an official ban on the drug. Severe measures were taken by the colonial government to suppress the khat trade – drivers of trucks used for its transport were jailed and their vehicles destroyed (Elmi 1983: 166). But prohibition proved futile and, in 1957, the protectorate replaced the ban on khat with an import tax on the leaf (Brooke 1960: 57).

After the two Somalilands joined as a republic in 1960, the use of khat rapidly gained popularity in the “Southern Regions” (the former Italian Somaliland), coinciding with increased urbanization and the improvement of surface and air transportation. For two decades the government attempted simply to discourage the practice and finally proscribed it in 1983.

Enforcement efforts, however, were not effective, and as a consequence of the civil war that began in 1991 and the continuing political turmoil, there are no legal restrictions on khat in Somalia at the present time. In fact, it is believed that strife in the capital, Mogadishu, is fueled by the struggle among clan leaders for control of lucrative khat imports and distribution. The value of khat, it is reported, far exceeds that of any other commodity that Somalia imports, including food and weapons (Randall 1993: 15).

In the northern administrative regions of Somalia, the chief market for khat is the city of Hargeysa. A small amount of khat is locally cultivated, but most is imported by trucks from the Harer area in Ethiopia.

Kenya. The preeminent khat-growing area of Kenya is the Nyambeni Hills of the Meru District, extending northeast from the foot of Mt. Kenya for about 120 km. Khat that is cultivated and sold under license by farmers of the Meru ethnic group is the main source of the leaf for consumers in Kenya and for export to Somalia, Uganda, Tanzania, Zaire, and Zambia (Matai 1983: 88).

From local market towns, khat is transported to Nairobi – a fast four-hour drive – or by air from Isiolo to northern market towns. The khat trade is highly seasonal. Shortages and high prices prevail at markets during the months from June to October, when new growth of leaves and shoots is at a minimum (Hjort 1974: 29).

Somalia is the main export market. Every day at dawn, some 15 to 20 airplanes, each with about a ton of khat stuffed in burlap bags, take off for Somalia from Nairobi International Airport. Other flights leave from the town of Meru to smaller markets for khat in Tanzania and Uganda (Lorch 1994: A8). Khat is also flown as cargo from Nairobi to Britain, where it is a legal import. London and Cardiff, where many of the dockworkers are Somalis, are major markets.

The Western Indian Ocean Region

Madagascar. The cultivation, sale, and consumption of khat are legal in Madagascar, where the drug is grown and consumed chiefly by members of the Antakarana ethnic group in the northern extremity of the island. Local names for the plant are “katy” and “gat” (Thomas Herlehy and Daniel Randiriamanalina, 1995, personal communication).

The introduction and early history of khat in Madagascar is obscure. It is possible that *C. edulis* was first brought to the island by Arab traders and immigrants several hundred years ago. Louis Molet (1967: 25) postulated that khat was introduced and spread by Yemeni Arabs and later by Muslims from the Comoro Archipelago. Traders from Kilwa and from Zanzibar may also have played a role in the introduction of khat to Madagascar. In any case, for generations, immigrants from Yemen and the Comoro Islands have settled among the Antakarana and Sakalava ethnic groups in northern Madagascar.

A minority of the population, commonly referred to as “Arabs,” are of mixed Yemeni, Comoran, and Malagasy ancestry. They profess Islam, affect Arab (Yemeni) dress, and are said to be foremost in the commercial production of khat (Sharp 1995, personal communication).

It is mostly men that use khat, but the number of women users is said to be increasing. The retail price of a bundle of khat is about 5,000 Malagasy francs (approximately \$1.25), which is expensive for most consumers (Sharp 1995, personal communication). As in Yemen, saliva and the juice of the leaves are swallowed, and the residue of chewed leaves is expectorated.

Methods of Khat Consumption

Khat as a Beverage

The practice of drinking water infusions and decoctions of khat is very old but today accounts for a smaller part of khat consumption than in the past. Fresh leaves for chewing, now widely available, provide users with much more psychostimulation than desiccated, boiled leaves. Nevertheless, khat prepared as a beverage is mildly stimulating, and some writers have extolled it as "an excellent beverage plant and worthy of exploitation" (Hill 1952: 481). "The leaf [has] a slightly bitter flavor with a strong, sweet taste of liquorice and has been regarded as nourishing" (Watt and Breyer-Brandwijk 1962: 179).

Catha edulis is one of many plants in Africa used to make "bush tea." In the Cape Province of South Africa, bush teas are popular as tonics and as treatments for urinary and digestive problems. J. M. Watt and M. G. Breyer-Brandwijk (1962: 590) have described two methods of preparing bush tea from young twigs and leaves of the khat tree. In one, flowering shoots are first fermented by being piled in a heap and then allowed to dry. In the other, the leaves are "sweated" in an oven before they are dried in the sun. According to Watt, well-prepared bush tea has a sweet aroma (Watt and Breyer-Brandwijk 1962: 590).

W. Cornwallis Harris, who traveled widely in Ethiopia during the years 1841 and 1842, reported that leaves of khat, well dried in the sun, were either chewed, boiled in milk, or infused in water. "By the addition of honey," wrote Harris, "a pleasant beverage is produced, which being bitter and stimulative, dispels sleep if used to excess" (1844: 423-4). In the region of Ifat (Welo Province), he found that the fresh leaves were chewed as an astringent medicine or taken to dispel sleep, "a decoction in water or milk being drunk as a beverage, which tastes bitter enough" (Harris 1844: 407). According to Simoons (1960: 115), Christians in Debra Tabor (Gondar Province, Ethiopia) use khat leaves to flavor mead (*tedj*). The addition of *tedj* to a water infusion of khat yields a brown, bitter, mildly intoxicating beverage (Schopen 1978: 85).

Khat as a Masticatory

The chewing of fresh leaves and stem tips is the most common way of using khat as a psychostimulant. Everywhere the technique is much the same. Derek Peters (1952: 36) describes how khat is used as a masticatory in Somaliland. From the user's supply of fresh khat, tender leaves and shoots are carefully selected and stripped from a few branches, compressed into a small mass, placed in the mouth, and chewed. Saliva and plant juices are swallowed. Copious amounts of cold water or other cold beverages are drunk. After 10 to 15 minutes of chewing, most of the juices are extracted from the wad. The process is repeated until - after 3 or 4 hours - the user's supply of khat is

exhausted. According to Weir (1985: 97), most consumers chew about 100 to 150 grams of picked leaves in one session; some chew as few as 50 grams, and a minority chew more than 200 grams.

In Yemen, the residue of chewed leaves in the mouth is not swallowed but compressed into a wad and stowed in one cheek (Kennedy 1987: 88). At the end of the chewing session it is expelled into a spittoon. In Ethiopia and East Africa, the chewed parts of the leaves are swallowed with the juice and saliva. Amare Getahun and A. D. Krikorian (1973: 371-2) have suggested that ingested residue may be an important part of the daily intake of food, especially in the case of heavy users of the drug. Other investigators suspect that tannin in the residue may be responsible for the gastrointestinal discomfort that is common among khat chewers.

Khat as a Paste

In Ethiopia, water and honey are added to crushed, dried leaves of khat, and the ingredients are worked into a paste. In Somalia, a paste is made of finely ground dried khat, water, sugar, cardamom, and cloves. Khat paste is commonly eaten by the elderly and by travelers, according to Schopen (1978: 84-5).

Other Modes of Use

Not uncommonly, elderly persons, if unable to chew effectively, pound fresh leaves of khat in a small mortar (which they carry on their person) and drink the juice (Brooke 1960: 53; Kennedy 1987: 88). In eighteenth-century Yemen, travelers and old persons used *madquq*, which is simply khat that is pressed and left to dry in a darkened room for seven days and then pulverized by mortar and pestle. A portion of the dry, granular khat is placed in the mouth, mixed with saliva, and swallowed (Schopen 1978: 86).

There are also brief references in the literature to the practice of smoking khat. For example, "The dried leaves [in Africa] are also sometimes smoked" (Margetts 1967: 358), and "In Arabia the leaves may be dried and smoked like tobacco" (Greenway 1947: 99). Y. Z. Hes mentions that in the towns of Yemen, there are places where one may go in order to make and smoke khat cigarettes (1970: 283-4). But today, with the widespread availability of the freshly cut product, this method of consumption appears to be little used.

Food and Nutritional Characteristics

Karsten Niebuhr, in his book on travels in Arabia published in 1792, ascribed to khat "the virtues of assisting digestion and fortifying the constitution against infectious diseases. Yet its insipid taste gives no indication of its extraordinary virtues" (Niebuhr 1792, Vol. 2: 353-4, cited by Krikorian 1984: 119).

The astringent effect of khat induces intense thirst, and in addition to water, such beverages as tea, coffee, and commercial colas are consumed during

the chew. Beer, both commercial and homemade, is a popular thirst quencher with chewers in Kenya and Somalia. Sometimes the host of a khat party in Yemen provides sherbet for the guests during a chewing session.

The nutritional value of khat is important because, to some unknown extent, the leaf is a substitute for conventional foods that would be eaten by the majority of users who chew khat regularly. Solid food is avoided during the chewing session, and for some hours after. Indeed, a well-known effect of khat is the temporary loss of appetite. "The laborers and shopkeepers take khat and regard it as the best substitute for food," write Mulugeta Assefa (1983: 74).

In a 1952 analysis of samples of miscellaneous tropical and subtropical plants and plant products by Margaret Mustard (1952: 31), khat was found to contain "exceptionally large amounts" of ascorbic acid (vitamin C) in leaves and branch tips, and "the ancient custom of qat chewing in Arabia may be inadvertently supplying the people of that country with some of their daily requirement of ascorbic acid" (1952: 34). It is interesting to note that ascorbic acid has been reported to act as an antidote to amphetaminelike substances, such as khat (Krikorian 1984: 154).

The Ethiopian Nutrition Survey (Khat [or chat] 1959: 169), in a nutritional analysis of samples of fresh khat leaves and tender stems, reported that in addition to its vitamin C content, "khat can contribute important amounts of niacin, beta-carotene [sic], calcium and iron to the user's diet" (See Table III.7.1).

According to Weir (1985: 42-3), the diet of khat chewers conforms to traditional Greco-Arab medical doctrine about the functioning of the human body. A healthy body is thought to be one in which an equilibrium is maintained among the four constituent humors - blood, phlegm, yellow bile, and black bile. Each humor has qualities drawn from the categories wet/dry and hot/cold. Blood is moist and hot, phlegm is damp and cold, yellow bile is dry and hot, and black bile is dry and cold. All substances that are taken internally, including khat, food, and medicines, are capable of upsetting the balance among the four humors by causing an excess or a diminution of their qualities when ingested. Khat is cooling and drying:

In order to achieve the most enjoyable results from an afternoon's qat party, and to offset the "cooling" effects of qat, a man should make himself as hot as possible by eating beforehand a good lunch composed of "hot" foods, and avoiding those classified as "cold." Hot foods are boiled mutton, mutton broth, fenugreek (*bilbah*) broth, chili pepper, white radish, wheat bread and sorghum porridge. . . . The cooling effects of qat are also counteracted during the qat party by closing all the windows of the room and generating a warm and stuffy atmosphere (Weir 1985: 43).

Weir (1985: 172) mentions that fenugreek is, next to mutton, the most important component of the midday meal, which is the main Yemeni meal. In Kennedy's opinion, there is traditional wisdom in the belief that fenugreek soup and a meat sauce gravy are essential components of a "greasy carbohydrate meal [that] nullifies at least some of the potentially negative side effects of the qat" (1987: 82).

Kennedy (1987: 81-2) also points out that it is essential that spicy and well-salted foods be eaten in order to heighten the thirst of the khat user during the chewing session. Pleasure in drinking cool water during the session is an important part of the chewing experience.

Medicinal Uses of Khat

Khat is included in only a few of the many Arab pharmacopoeias and has been very little used in formal medicine. As might be supposed, however, it has a place in folk medicine and traditional remedies. For example, according to Peter Merab, among the dervishes and other Muslim holy orders, chewed khat is spat upon patients as a preliminary step in treatment of the sick. He also mentions that infusions of khat are drunk to treat illness (1921: 176).

In Somalia, khat is believed to have curative properties in treating urinary problems and gonorrhea (Peters 1952: 36). In Tanzania, coughing, asthma, and other respiratory problems are dealt with by drinking khat infusions, and pieces of the tree root are eaten to relieve abdominal pains (Greenway 1947: 99).

In the eighteenth century, Yemenis believed that the use of khat prevented epidemics and that, wherever the khat plant grew, pestilence would not appear (Schopen 1978: 87-8). Khat is used as an analgesic in Madagascar, where "teas" of many different plants are consumed for medicinal purposes. Tea from khat is used by elderly persons to relieve rheumatic ache and by women to ease the pain of menstrual cramps (Sharp 1995, personal communication).

The Psychophysiological Effects of Khat

The psychotropic effects of khat first drew international attention in 1935, when two technical papers on the plant were discussed in the League of Nations' Advisory Committee on the Traffic in Opium and Other Dangerous Drugs (*Bulletin on Narcotics* 1980: 1).

For many years, it was believed that the alkaloid cathine, *d*-norpseudoephedrine, was the main stimulant in khat. But in 1975, a laboratory study - using samples of fresh khat (instead of dried leaves as had been used in previous work) - resulted in the detection and isolation of cathinone, a phenylalkylamine characterized as (-)- α -aminopropiophenone (Szendrei 1980: 5). Cathinone, 8 to 10 times stronger than cathine in its effect, is the predominant psychotropic

agent in khat (Kalix 1994: 71), which is the only plant known to produce cathinone (LeBelle 1993: 54).

The transformation of cathinone into other less-potent products, such as cathine, begins shortly after the leaves and shoots are removed from the tree (Kalix 1994: 72). With a plasma half-life of only 1.5 hours, cathinone's potency is lost within 48 to 72 hours after harvesting (Kalix 1994: 72), which is

the reason khat users prefer to chew only fresh leaves and shoots of the plant. However, deep-freezing prolongs its potency for months (Drake 1988: 532-3).

Alkaloids from khat, used in patent medicines manufactured in Europe during the 1920s and 1930s, were derived only from dried leaf material and, therefore, contained little or no cathinone.

Table III.7.1. *Nutritional components of khat (Catha edulis)*

Vitamins and minerals	Sugars and sugar alcohols	Miscellaneous substances
Ash: 11.59 gm/100 gm dried leaves (includes Mg, Fe, Ca, Cl)	Mannitol	Rubberlike substances
Ascorbic acid: .136 gm/100 gm leaves and twigs 0.324 gm/100 gm leaves only	-	-
-	-	Rubberlike substances: volatile oil (traces)
Gm/100 gm plant material Ash 1.6% Ascorbic acid 0.161 Thiamine <0.00005 Niacin 0.0148 Riboflavin <0.00005 β -carotene 0.0018 Iron 0.0185 Calcium 0.290	-	Protein: 5.2% Fiber: 2.7%
-	-	Free amino acids, aspartic acid; threonine; serine; glutamic acid; proline, glycine; alanine; valine; leucine; isoleucine; tyrosine; α -aminobutyric acid; histidine; ornithine; arginine; tryptophan; phenylalanine.
-	Reducing sugars: as hexose: 1.4 gm/100 gm plant material (portion of total sugars may be galactose); dulcitol	Free amino acids: lysine; α -aminobutyric acid; phenylserine; phenylalanine choline (0.05%) (dried plant)
-	Glucose, fructose; rhamnose (free); xylose, galactose; dulcitol	-
-	-	Flavonoids; glycosides or carbohydrates; volatile oils
-	-	Steam distillate of dried leaves and young twigs (of Yemen origin) yielded 0.04-0.08% of a yellow essential oil. There were 5 hydrocarbons (α and β -pinens, terpinolene, β -phellandrene, and ocimene); and 6 oxygenated terpenes (α and β -thujone, fenchone, linalool, α -terpineol, and nerol.

Source: After Krikorian and Getahun (1973).

The Khat Experience

A subjective account of khat-induced psychophysiological effects is given by Kennedy. The account is derived from his own experimentation with the drug and from interviews with 803 Yemeni users of khat:

The first experiential effects of qat are a gradually developing mild euphoria, i.e., alertness, feelings of contentment, confidence, gregariousness, and the flowing of ideas, called the *kayf*. ["*Kayf*" is a general term in Arabic for a state of pleasurable well-being]. Under certain variable conditions such as amount of the drug, food intake, personality and physical condition, this *kayf* phase occasionally develops into confusion. Ordinarily however, the experience gradually transforms into an "introvert phase" of quiet contemplation in which the stimulation of thought processes continues at a rapid pace while physical and vocal activities diminish. Since this is often accompanied by fatigue, previous problems or pessimistic personality characteristics may transform this phase into a mild temporary lowering of mood.

The postsession effects of qat continue for several hours, and, again depending upon the same variable conditions, many of them are rather negative. Sleep is inhibited, hunger is diminished, sexual desire and performance may positively or negatively be affected, and irritability may occur. However, often people experience positive aftermaths, such as feeling closer to Allah, more understanding, and a desire to work. When we look at our interview data pertaining to the experiential aspects of qat-use we find consistent male-female differences, with more males generally reporting the experiences. In the eyes of the average chewer the rewarding experiences associated with qat chewing outweigh the unrewarding ones; they are much more regular and predictable (Kennedy 1987: 130-1).

Adverse Effects of Khat

An apocryphal account of khat's introduction into Yemen from Ethiopia is given by the historian and traveler, Ibn Fadl Allah al'Umari in his work, *Masalik al-Absar*, written between A.D. 1342 and 1349. The king of Yemen, curious about khat, asked an advisor, an Ethiopian familiar with the drug, to describe the effects of chewing the leaf:

Upon learning that it virtually eliminated the desire to eat, to drink [alcohol?], and to have sexual intercourse, [the king] told him: "And what other pleasures are there in this base world besides those? By God, I will not eat it at all; I only spend my efforts on those three things; how am I going to use such a thing

which will deprive me of the pleasures that I get from them?" (Krikorian 1984: 26-7).

Beyond question, a temporary loss of appetite is effected by khat chewing. In laboratory tests, cathinone and cathine markedly inhibit the food intake of rats (*Bulletin on Narcotics* 1980: 84). In addition, nearly all users experience difficulty in sleep following the use of the drug. Only daily heavy chewers reported that their sleep was not inhibited by khat (Kennedy 1987: 128). There is also little doubt that sexual experience is affected by khat, but opinions vary as to whether sexual desire is heightened and ability to perform is enhanced, or whether there is a loss of libido. Examples of both effects are found in the literature: "While sex interest is heightened at first, depressed *libido* and *potentia sexualis* come as the drug effect is maintained, and chronic users may develop impotence" (Margetts 1967: 359; see also Krikorian 1984: 151).

A negative view of the drug, based on its anaphrodisiac properties, is paradoxical, however, because relevant demographic characteristics of the chief khat-using countries suggest that "depressed libido" is hardly a problem. Yemen, for example, has a total fertility rate (the average number of children a woman will have throughout her childbearing years) of 7.7, which is the world's highest, save for Gaza's 8.1 (Population Reference Bureau 1995).

Perhaps the most common ailments related to the level of khat use are gastritis (acute and chronic) and constipation (Kennedy 1987: 218). Tannins in khat (chiefly flavenoids) have long been suspected of adversely affecting the gastric system. At the present time, it is known that tannin content is high, probably between 5 and 15 percent (Kennedy 1987: 185). Many of the most expensive kinds of khat have reddish tints to leaves and shoots, and these kinds are purported to have a lower content of tannins than the others (Drake 1988: 532).

Among women, there is a strong association of the level of khat use with liver and urinary problems. Khat chewing may also be a maternal practice harmful to the fetus. It has been found that healthy full-term infants have a significantly lower average birth weight if the mothers were either occasional or habitual khat chewers (Eriksson, Ghani, and Kristiansson 1991: 106-11).

Legal Considerations

Since the 1950s, khat has been on various agendas of meetings under the auspices of the United Nations Commission on Narcotic Drugs and of the World Health Organization. For many years, the attention of these organizations was directed to the question of agreeing upon a pharmacological definition of *C. edulis*. Eventually, in 1973, it was placed in a group of "dependence-producing drugs," although there is a

lack of evidence that khat chewing produces "dependence" in the usual sense of the word.

At the present time, 137 countries are parties to at least one of the three Conventions (1961, 1971, and 1980) of the United Nations International Drug Control Programme. Because khat contains the alkaloids cathinone and cathine, it is a scheduled drug on the 1971 Convention's List of Psychotropic Substances. Parties to the Convention agreed to prohibit the entry of khat without an import permit. However, most parties to the U.N. conventions do not strictly enforce the control measures for khat as required by the treaties.

Europe

The Scandinavian countries are among the minority of European nations that strictly enforce a ban on the import and export of khat. Consumers in the United Kingdom and Germany may purchase khat legally, airborne from Ethiopia, Kenya, and, recently, Yemen (Rushby 1995: 16). In Rome, where use of the drug is popular among Somali expatriates, it is rumored that airport customs inspectors of flights arriving from Ethiopia, Kenya, Yemen, and Israel seldom inquire closely as to the nature of fresh bundles of stems and leaves identified on import declaration forms as "floral stuff" (Nencini et al. 1989: 257).

The United States

Under United States federal law, it is illegal to possess khat for personal use or to import, cultivate, dispense, or distribute it. The alkaloid cathinone is placed in Schedule I, subsection (f) Stimulants, of the Controlled Substances Act (CSA). The CSA defines Schedule I drugs as "highly abusable substances . . . which have no currently accepted medical use in the United States" and which may be used lawfully only in research situations authorized by the Drug Enforcement Administration (U.S. Department of Justice 1988: 4). As provided by the CSA, the penalty for simple possession of khat is imprisonment of not more than one year and a minimum fine of \$1,000, or both (Code of Federal Regulations, 21 1994: 1308.12, 841[a], 841[b], 844[a]).

It is believed that most khat enters the United States from Canada in small quantities concealed in luggage aboard aircraft or in private vehicles. Seizures of khat have been made at international airports in Champlain, N.Y., New York City, Newark, Chicago, Dallas, and St. Louis. A total of about 800 kilograms of khat was seized by the U.S. Customs Service between August 1991 and October 1992. In March 1993, the retail price of a kilogram of khat ranged from 30 to 60 dollars (U.S. Department of Justice 1993: 2).

Canada

Khat is currently classified as a "New Drug" in Canada, that is, "new" in the sense that "it has not been sold in Canada for sufficient time to establish the safety and

effectiveness of the substance for use as a drug" (Langlois 1995, personal communication). As a New Drug, the sale of khat is prohibited, but not its possession import, export, or consumption. New legislation would place fresh leaves of khat in Schedule II and confine import, export, cultivation, sale, and possession to medical and scientific purposes. Importation would require an official permit. Violations would be subject to fines not to exceed 5,000 dollars and or imprisonment not exceeding three years (Langlois 1995, personal communication).

Clarke Brooke

The author acknowledges the valuable assistance of the following individuals in preparing this paper: P. K. Bailey, Badalkhan Baloch, N. A. Baloch, Richard Halse, W. W. Heiser, Ross Hossie, C. C. Held, Erich Isaac, Fritz Kramer, John Mandeville, Martin Novack, Thomas Poulsen, Jean Pryor, Ziva Razafintsalama, Hiromi L. Sakata, Lesley Sharp, Arnon Soffer, and Frank Wesley.

Bibliography

- Al-Hammad, Abed-el-Kader Shiba. 1983. The Shari'ah's position on khat. In *The health and socio-economic aspects of khat use*, ed. B. Shahandeh, R. Geadah, A. Tongue, E. Tongue, et al., 225-30. Lausanne.
- Al-Meshal, I. Ibrahim, Shueb Qureshi, Abdul Rehman, and Mohammed Tariq. 1991. The toxicity of *Catha edulis* (khat) in mice. *Journal of Substance Abuse* 3: 107-15.
- Al-Thani, I. M. 1983. Development: The Saudi solution for the problem of khat. In *The health and socio-economic aspects of khat use*, ed. B. Shahandeh et al., 181-94. Lausanne.
- Assefa, Mulugeta. 1983. Socio-economic aspects of khat in the Harrarghe administrative region (Ethiopia). In *The health and socio-economic aspects of khat use*, ed. B. Shahandeh et al., 72-7. Lausanne.
- Azaïs, R. P., and R. Chambard. 1931. *Cinq années de recherche archéologiques en Éthiopie. Province du Harar et Éthiopie méridionale*. Paris.
- Baird, D. A. 1951. Indian remedies for poor memories. *British Medical Journal* 2: 1522.
- Balint, G., H. Ghebrekidan, and E. Balint. 1991. *Catha edulis*, an international socio-medico problem with considerable pharmacological implications. *East African Medical Journal* 68: 555-61.
- Bally, P. R. O. 1945. *Catha edulis*. *East African Medical Journal* 22: 2-3.
- Bill C-7. First session, Thirty-fifth Parliament, 42 Elizabeth II, 1994. Ministry of Health, House of Commons, Canada. Ottawa.
- Brooke, Clarke. 1958. The durra complex of the Central Highlands of Ethiopia. *Economic Botany* 12: 192-204.
- 1960. Khat (*Catha edulis*): Its production and trade in the Middle East. *Geographical Journal* 126: 52-9.
- Browne, Dallas. 1990. Qat use in New York City. Research Monograph 105, 1990: 464-5. National Institute on

- Drug Abuse. U.S. Department of Health and Human Services. Washington, D.C.
- Bulletin on Narcotics*. 1980. Special issue devoted to *Catha edulis* (khat), Vol. 32, No. 3.
- Burton, Richard F. 1910. *First footsteps in East Africa*. London.
- Chevalier, A. 1949. Les cat's d'Abyssinie et d'Afrique Orientale. *Revue de Botanique Appliquée et d'Agriculture Tropicale* 29: 413-17.
- Code of Federal Regulations: Food and Drugs*. 21. 1994. Part 1300 to End revised as of April 1, 1994. Office of the Federal Register, National Archives and Records Administration. Washington, D.C.
- Drake, P. H. 1988. Khat-chewing in the Near East. *The Lancet* 1 (March 5): 532-3.
- Duke, James. 1985. *CRC handbook of medicinal herbs*. Boca Raton, Fla.
- Economist Intelligence Unit Limited, The. 1987. *Djibouti EIU Country Report No.4, 1987*. London.
1988. *Djibouti EIU Country Report No.1, 1988*. London.
1989. *Djibouti EIU Country Profile 1989-90*. London.
1994. *Ethiopia EIU Country Report No.1, 1994*. London.
1989. *Yemen EIU Country Report 1989*: 56. London.
1992. *Yemen EIU Country Report No. 3, 1992*: 21-9. London.
1994. *Yemen EIU Country Report 1994*: 19-27. London.
1994. *Yemen EIU Country Profile 1994-95*. London.
- Elmi, Abdullahi. 1983. The chewing of khat in Somalia. *Journal of Ethnopharmacology* 8: 163-76.
- Elmi, A. S., T. H. Ahmed, and M. S. Samatar. 1987. Experience in the control of khat-chewing in Somalia. *Bulletin on Narcotics* 39: 51-7.
- Eriksson, M., N. A. Ghani, and B. Kristiansson. 1991. Khat-chewing during pregnancy - effect upon the off-spring and some characteristics of the chewers. *East African Medical Journal*: 106-11.
- Geissshusler, S., and R. Breneissen. 1987. The content of psychoactive phenylpropyl and phenylpentenyl khatamines in *Catha edulis* Forsk. of different origin. *Journal of Ethnopharmacology* 19: 269-77.
- Getahun, Amare, and A. D. Krikorian. 1973. Chat: Coffee's rival from Harar, Ethiopia. I. Botany, cultivation and use. *Economic Botany* 27: 353-77.
- Giannini, A. J., N. S. Miller, and C. E. Turner. 1992. Treatment of khat addiction. *Journal of Substance Abuse Treatment* 9: 379-82.
- Gidron, Avner. 1993. Khat or not? *World Press Review* 40: 30.
- Great Britain, Naval Intelligence Division. 1946. *Western Arabia and the Red Sea*. Geographical Handbook Series, B.R. 527. Oxford.
- Greenway, P. J. 1947. Khat. *East African Agricultural Journal*: 98-102.
- Harris, W. Cornwallis. 1844. *The highlands of Aethiopia*, Vol. 3. Second edition. London.
- Hes, Y. Z. 1970. Some notes concerning the use of kat (*Catha edulis*). Ha-Refu'ah: 283-4 (in Hebrew). Jerusalem.
- Hess, J. J., 1927. Kat. In *The encyclopedia of Islam*, Vol. 2, ed. M. Th. Houtsama, A. J. Wensinck, T. W. Arnold, et al., 808. Leiden.
- High on gat. *Jerusalem Post* 1996. International Edition. November 10-16, pp. 16-17.
- Hill, A. F. 1952. *Economic botany*. New York
- Hill, Bob. 1965. Cat (*Catha edulis* Forsk). *Journal of Ethiopian Studies* 3: 13-23.
- Hill, C. H., and A. Gibson. 1987. The oral and dental effects of q'at chewing. *Oral Surgery, Oral Medicine, Oral Pathology* 63: 433-6.
- Hjort, Anders. 1974. Trading miraa: From school-leaver to shop-owner in Kenya. *Ethnos* 39: 27-43.
- Ingrams, Harold. 1966. *Arabia and the Isles*. Third edition. London.
- Islam, M. W., O. A. Al-Shabanah, M. M. Al-Harbi, and N. M. A. Al-Gharbly. 1994. Evaluation of teratogenic potential of khat (*Catha edulis* Forsk). in rats. *Drug and Chemical Toxicology* 17: 51-68.
- Kafih, J. 1963. *Halikbot Teman* (Folkways of Yemen, Jewish life in San'a). Jerusalem.
- Kalix, Peter. 1990. Pharmacological properties of the stimulant khat. *Pharmacological Therapy* 48: 397-416.
1994. Khat an amphetamine-like stimulant. *Journal of Psychoactive Drugs* 23: 69-74.
- Kalix, Peter, and Olav Braenden. 1985. Pharmacological aspects of the chewing of khat leaves. *Pharmacological Reviews* 37: 149-64.
- Karawya M. S., M. A. Elkicy, and M. G. Ghourab. 1968. A study of *Catha edulis* Forsk. growing in Egypt. *Journal of the Pharmaceutical Sciences of the United Arab Republic* 9: 147-57.
- Kennedy, John G. 1987. *The flower of paradise*. Dordrecht, Netherlands.
- Khat. 1956. *Bulletin on Narcotics* 8: 6-13.
- Khat (or chat). 1959. *Ethiopia Nutrition Survey. A Report by the Interdepartmental Committee on Nutrition for National Defense* 30: 166-9. Washington, D.C.
- Krikorian, A. D. 1984. Kat and its use: An historical perspective. *Journal of Ethnopharmacology* 12: 115-78.
- Krikorian, A. D., and Amare Getahun. 1973. Chat: Coffee's rival from Harar, Ethiopia. II. Chemical composition. *Economic Botany* 27: 378-89.
- Langdale-Brown, I., H. A. Ostmasten, and J. G. Wilson. 1964. *The vegetation of Uganda and its bearing on land-use*. New York.
- LeBelle, M. J. 1993. Gas chromatographic-mass spectrometric identification of chiral derivatives of khat. *Forensic Science International* 61: 53-64.
- Lorch, Donatella. 1994. Despite war, famine and pestilence, the khat trade thrives in East Africa. *New York Times*. December 14, p. A8.
- Mack, Ronald. 1995. Khat on a hot tin roof. *North Carolina Medical Journal* 56: 112-14.
- Margetts, Edward. 1967. Miraa and myrrh in East Africa - clinical notes about *Catha edulis*. *Economic Botany* 21: 358-62.
- Matai, C. K. 1983. Country report - Kenya. In *The health and socio-economic aspects of khat use*, ed. B. Shahandeh et al., 87-90. Lausanne.
- McKee, C. M. 1987. Medical and social aspects of qat in Yemen: A review. *Journal of the Royal Society of Medicine* 80: 762-5.
- Merab, Peter. 1921, 1922, 1923. *Impressions d'Éthiopie (L'Abyssinie sous Menelik II)*. 3 vols. Paris.
- Molet, Louis. 1967. Cadres pour une ethnopsychiatrie de Madagascar. *L'Homme* 7: 5-29.
- Morghem, M. M., and M. I. Rufat. 1983. Cultivation and chewing of khat in the Yemen Arab Republic. In *The health and socio-economic aspects of khat use*, ed. B. Shahandeh et al., 215-16.
- Murdock, George P. 1959. *Africa: Its peoples and their culture history*. New York.
- Mustard, Margaret. 1952. Ascorbic acid content of some miscellaneous tropical and subtropical plants and plant products. *Food Research* 17: 31-5.
- Nencini, Paolo, Abdullahi Mohammed Ahmed, and Abdullahi Sheik Elmi. 1986. Subjective effects of khat chewing in humans. *Drug and Alcohol Dependence* 18: 97-105.

- Nencini, Paolo, Maria Caterina Grassi, Abdikadar Ashkir Botan, et al. 1989. Khat chewing spread to the Somali community in Rome. *Drug and Alcohol Dependence* 23: 255-8.
- Nyrop, Richard, ed. 1985. *The Yemens: Country studies*. 2d edition. Washington, D.C.
- Ofcansky, Thomas, and LaVerle Berry, eds. 1993. *Ethiopia: A country study*. Fourth edition. Washington, D.C.
- Owen, P. J. 1910. *Catha edulis*. *Journal of the Society of Chemical Industry* 29: 1091.
- Pankhurst, Richard. 1961. *An introduction to the economic history of Ethiopia*. London.
- Pantelis, Christos, Charles G. Hindler, and John C. Taylor. 1989. Use and abuse of khat (*Catha edulis*): A review of the distribution, pharmacology, side effects, and a description of psychosis attributed to khat chewing. *Psychological Medicine* 19: 657-68.
- Paris, R., and H. Moyse. 1958. Abyssinian tea (*Catha edulis* Forsk., *Celastraceae*). A study of some samples of varying geographical origin. *Bulletin on Narcotics* 10: 29-34.
- Peters, Derek. 1952. Khat: Its history, botany, chemistry and toxicology. *The Pharmaceutical Journal* 169: 17-18, 36-7.
- Population Reference Bureau. 1995. *World Population Data Sheet: 1995*. Washington, D.C.
- Randall, Teri. 1993. Khat abuse fuels Somali conflict, drains economy. *Journal of the American Medical Society* 269: 12-15.
- Revri, Raman. 1983. *Catha edulis* Forsk. *Geographical dispersal, botanical, ecological and agronomical aspects with special reference to Yemen Arab Republic*. Göttingen.
- Rosenzweig, Kurt, and Patricia Smith. 1966. Peridontal health in various ethnic groups in Israel. *Journal of Periodontology* 1: 250-9.
- Rushby, Kevin. 1995. The high life. *Geographical* 67: 14-17.
- Santagata, Fernando. 1940. *L'Harar. Territorio di pace e di civiltà*. Milan.
- Sayer, J. A., C. S. Harcourt, and N. M. Collins, eds. 1992. *Conservation of tropical forests*. Africa and London.
- Schopen, Armin. 1978. *Das Qat: Geschichte und Gebrauch des Genussmittels Catha Edulis Forsk in der Arabischen Republik Jemen*. Wiesbaden.
- Scott, Hugh. 1947. *In High Yemen*. London.
- Seyoum, E., Y. Kidane, H. Gebru, and G. Sevenhuysen. 1986. Preliminary study of income and nutritional status indicators in two Ethiopian villages. *Food and Nutrition Bulletin* 8: 37-41.
- Shahandeh, B., R. Geadah, A. Tongue, et al, eds. 1983. *The health and socio-economic aspects of khat use*. Lausanne.
- Simoons, Frederick. 1960. *Northwest Ethiopia: Peoples and economy*. Madison, Wis.
- Sullum, Jacob. 1993. Khat calls. *Reason* 24: 42-3.
- Szendrei, K. 1980. The chemistry of khat. *Bulletin on Narcotics*. 1980. *Special issue devoted to Catha edulis (khat)* 3: 5-36.
- Trimingham, J. Spencer. 1952. *Islam in Ethiopia*. London.
- Tutwiler, Richard, and Sheila Caprisco. 1981. *Yemeni agriculture and economic change: Case studies of two highland regions*. San'a.
- Uphof, George. 1974. *A dictionary of economic plants*. Würzburg.
- U.S. Department of Justice. 1988. *Drugs of abuse*. Drug Enforcement Administration. Washington D.C.
1993. *Khat Factsheet*. Drug Enforcement Administration. Washington, D.C.
1993. *Methcathinone (Cat) Factsheet*. Drug Enforcement Administration. Washington, D.C.
- Varisco, Daniel. 1986. On the meaning of chewing: The significance of qat (*Catha edulis*) in the Yemen Arab Republic. *International Journal of Middle East Studies* 18: 1-13.
- Vavilov, N. I. 1951. *The origin, variation, immunity and breeding of cultivated plants*, trans. K. Starr Chester. New York.
- Watt, J. M., and M. G. Breyer-Brandwijk. 1962. *The medicinal and poisonous plants of Southern and Eastern Africa*. Second edition. Livingston, Edinburgh, and London.
- Weir, Shelagh. 1983. Economic aspects of the qat industry in North-West Yemen. In *Economy and society in contemporary Yemen*, ed. G. R. Pridham, 64-82. Exeter, England.
1985. *Qat in Yemen: Consumption and social change*. London.
- Westphal, E. 1975. *Agricultural systems in Ethiopia*. In collaboration with J. M. C. Westphal-Stevens. Wageningen, Netherlands.
- Widler, Peter, Karoline Mathys, Rudolf Brenneisen, et al. 1994. Pharmacodynamics and pharmacokinetics of khat: A controlled study. *Clinical Pharmacology and Therapeutics* 55: 556-62.
- Zein, Ahmed Z. 1988. Polydrug abuse among Ethiopian university students with particular reference to khat (*Catha edulis*). *Journal of Tropical Medicine and Hygiene* 91: 71-5.

III.8 Kola Nut

Kola nut is an important stimulant and masticatory in Africa. It is about the size of a walnut or a chestnut and varies in color from dark red to creamy white (Chevalier and Perrot 1911; Cohen 1966; Agiri 1972; Lovejoy 1977-8, 1980). This fruit of the kola tree grows in pods that contain from 3 to 15 or more nuts. Pink and white nuts are generally valued highest because they have a sweeter taste and a greater caffeine content than other kola nuts. Kola is richer in caffeine than coffee and most teas and has 3 times as much starch as cacao (Heckel and Schlagdenhauffen 1884; Kraemer 1910; Chevalier and Perrot 1911; Pratt and Youngken 1956; Ramstad 1959). Known by various African names - *goro* or *gourou*, *ombéné*, *nangoué*, *kokkorokou*, and *matrasa* - kola is known as a heart stimulant because it contains kolanin along with caffeine, traces of the alkaloid theobromine (which exists in cacao as well), glucose, and strychnine (Table III.8.1). Both caffeine and theobromine stimulate the nervous system and the skeletal muscles, making kola a psychoactive substance (Kennedy 1987; Jones 1995).

Table III.8.1. *Chemical composition of the pod husk, testa, and nut of kola (dry matter %)*

Composition	Pod husk*	Testa*	Nut*
Crude protein (NS × 6.25)	10.22 (7.69–12.50)	14.00 (12.25–17.00)	8.06 (7.25–9.81)
Ash	6.05 (5.75–7.58)	11.00 (9.23–13.11)	2.55 (1.95–3.08)
Crude fiber	16.43 (15.72–18.26)	14.36 (13.33–15.77)	2.18 (1.83–2.57)
Ether extract	1.0 (0.8–1.4)	1.37 (0.94–1.56)	0.98 (0.96–1.0)
Nitrogen free extract	66.30 (61.52–70.04)	59.00 (52.56–64.25)	86.23 (83.54–88.01)
Caffeine	trace	0.44 (0.32–0.48)	1.58 (0.8–3.0)
Pectin	10.0 (9.24–11.57)	5.00 (4.54–5.50)	–
Total Polyphenols	8.80 (8.24–9.8)	3.63 (2.50–4.20)	6.7 (5.6–8.5)
Total nitrogen	1.63 (1.23–2.0)	2.24 (1.96–2.72)	1.28 (1.16–1.56)
Nonprotein nitrogen	0.69 (0.5–0.93)	0.34 (0.30–0.82)	0.88 (0.76–1.16)

*The diagram above shows the chemicals that are contained in kola nuts, which make kola such an important masticatory in Africa.

Source: Ogutuga (1975), pp. 121–5.



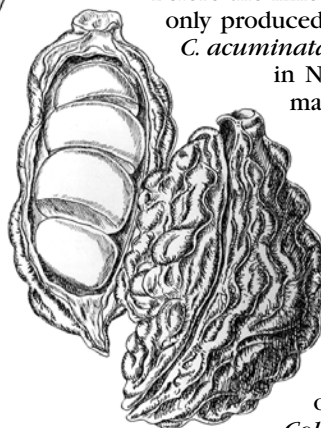
Kola

Kola Production

The genus *Cola* is of tropical African origin (Freeman 1893; Irvine 1948; Cohen 1966; Dickson 1969) and belongs to the *Sterculiaceae* family, the species of which are most abundant in tropical Asia (Oliver 1868; Heckel and Schlagdenhauffen 1884). A number of different species grow in the region between Sierra Leone and the Congo (Heckel and Schlagdenhauffen 1884; Kreiger 1954; Kola 1957; Dickson 1969; Morgan and Pugh 1969; Brooks 1980; Anquandah 1982). But of some 40 varieties, only *Cola nitida* and *Cola acuminata* are well known and widely used. The former was the kola of long-distance trade, whereas the latter was grown mainly for local consumption in the forest regions of Africa.

Before the nineteenth century, *C. nitida* was only produced west of the Volta River, and *C. acuminata* grew to the east, especially in Nigeria (Lovejoy 1980; Goodman, Lovejoy, and Sheratt 1995).

In the nineteenth century, however, the cultivation of *Cola anomala* and *Cola ballayi* had spread to southern Cameroon, with the nuts exported to the savanna region (Goodman et al. 1995). In Ghana, in addition to *C. nitida* and *C. acuminata*, there are other varieties that include *Cola cardifolia*, *Cola togoensis*



(Engl & Krause), *Cola johnsonii* (Stapf.), and *Cola verticillata* (Stapf. & Chev.) (Irvine 1969).

Although kola is indigenous to the forest zone of West Africa, the plant has traveled widely. It can be found in East Africa around Lake Nyanza, along the north coast of South America, and in portions of South Asia. In large part, this is because of European colonization. The British, for example, introduced kola into the East Indies, the Seychelles, Ceylon (Sri Lanka), Demerara (Guyana), Dominica, Mauritius, and Zanzibar. Similarly, the French introduced kola into Martinique, Guadeloupe, and Cayenne (Heckel and Schlagdenhauffen 1884).

Cultivation

Kola does well in forest soils and also in savanna areas of forest outliers so long as there is adequate moisture in the rooting zone and the soil has a high content of organic matter. The tree likes low elevations, deep, well-drained soils, and moderate rainfall. Good drainage is essential because the plant cannot withstand flooding (AERLS *Guide* 1982). It grows best in rich, deep soil in forest zones where rainfall does not exceed 50 to 60 inches annually – conditions characteristic of areas where the kola industry has persisted on a commercial scale (Heckel and Schlagdenhauffen 1884; Miles 1931).

The kola tree also flourishes in moist lands either at sea level or a little above. In Labogie in Nupe (Nigeria), kola plantations are situated in sheltered valleys at an elevation of 450 to 550 feet above sea level, where, according to a Royal Botanical Garden bulletin, “[t]he soil is a deep, black, sandy loam and is kept in a continuous state of moisture by streams that are in the valley” (1906: 89). The annual rainfall in this region is about 40 to 50 inches. Similarly, in southwestern Nigeria, kola is generally found in the better-drained areas with light, loamy soils of good depth (Russell 1955; Agiri 1972).

Germination

Polyembryo (in which more than one embryo results from a single ovule) occurs in *C. nitida* (Bodard 1954) and *C. acuminata*, along with multiple shoot/root production and adnation of auxiliary shoots to the main stem axis (Oladokun and Adepipe 1989). In fact, splitting or cutting the nuts brings quicker germination, although this reduces nut size, as well as initial growth after germination is slowed down. In addition, some of the cotyledons produce multiple roots and shoots, whereas others produce roots with no shoots (Oladokun and Adepipe 1989; Brown 1970).

High temperatures and light seem to have no effect on kola germination. Seeds that are freshly harvested take 3 to 9 months to germinate, although seeds stored for about 7 months usually do so within 3 or 4 months of sowing. There is also a pronounced difference between the germination patterns of stored and fresh kola seeds, with seedlings from

stored seeds tending to grow faster, larger, and more vigorously than those from fresh seeds (Ashiru 1969; Brown 1970; Karikari 1973).

Propagation

The propagation of kola nuts takes place after the land is prepared at the beginning of the rainy season. The farmer initially clears a portion of the forest and scrapes the soil into “hills” that are “beds” for yam, cassava, and cocoyam. Kola seeds are planted between these beds and spaced about 20 to 27 feet apart so that the food crops will provide shade for the young seedlings (Ashiru 1969; Brown 1970; Karikari 1973).

A kola tree bears fruit after 4 or 5 years and reaches maturity by the tenth year. Because it is a tree of the tropical forest, it is useful to interplant with food crops or with shade trees, such as coffee and cacao. Many of the kola groves or plantations in Ghana are found growing together with cacao (Ossei 1963). Because growth is very slow (about 3 meters in 4 years), the use of seed, rooting of cuttings, grafting of shoots on suitable rootstocks, budding, and aerial layering – or marcottage – are all techniques that have been employed to make seedlings available to farmers who have embraced the crop.

If kola is to be propagated from the seed, the nuts are taken from the pod and wrapped in plantain leaves. Then they are buried in the ground and watered every day until they sprout, after which the young seedlings are transplanted to bear fruit some 4 to 5 years hence (N.A.G. [Kumasi] 1905; Quarcoo 1969; AERLS *Guide* 1982; Szolnoki 1985). Seedlings are also raised in bamboo pots before being transplanted in the field. In some cases, bananas and plantains provide shade for the young plants (Ossei 1963).

Kola is also planted by one or another of four methods of vegetative reproduction (Pyke 1934). In the first, a branch of a healthy kola tree is cut into 2- to 3-foot-long pieces, which are then planted. After germination, these bear fruit in 2 or 3 years (N.A.G. [Kumasi] 1905; AERLS *Guide* 1982). Plant cuttings, one of the most frequently used present-day techniques in cloning fruit trees, is also one of the most ancient methods of vegetative propagation (Archibald 1955; Jones 1968; Ashiru and Quarcoo 1971; Ibikunle 1972; Ashiru 1975). It is very important that cloning be done only from high-yielding kola trees and, more importantly, from trees with a high degree of self- and cross-compatibility.

A second method of vegetative propagation is grafting. Various methods of grafting have been used in the production of kola clones (Ashiru and Quarcoo 1971; Gnanaratum 1972), with wedge grafting, saddle grafting, and the splice and whip-and-tongue grafts said to yield the best results (Jones 1968). In Nigeria, the side-and-wedge grafting technique is widely used.

Budding is a third method of vegetative propagation of kola. Nursery budding ensures the production of vigorous budlings and offers the possibility of virtu-

ally all of them continuing to bud in the field. By contrast, *in situ* budding cannot guarantee that 100 percent of the buds will take, although these usually do grow faster than those transferred from the nursery (Archibald 1955; Are 1965; Ashiru and Quarcoo 1971; Ibikunle 1972; Ashiru 1975).

Aerial layering (also known as marcottage) is the fourth method of vegetative propagation by which “an intact branch of a kola tree is induced to produce roots before it is severed from the mother tree” (Ashiru and Quarcoo 1971). After being severed, the cutting may be raised in a nursery or planted directly in the field. Branches that are growing vertically, or nearly so, are the best types to select for marcotting, and such branches should be between 2 and 3 years old. As a method of kola propagation, marcottage is old and well known, but it is also very cumbersome (Toxopeus and Okololo 1967).

Kola is – or at least was – sometimes planted on special occasions. In some cultures of sub-Saharan Africa, for example, it was the custom to bury the umbilical cord of a newborn baby with a kola seed. This indicated a safe delivery, and the kola tree that subsequently grew up became the property of the child. Cultural taboos also play a role in kola planting: In some parts of West Africa, a belief (perhaps spread to limit competition) that “he who plants kola will die as soon as the plant has flowered” militated against its cultivation. Instead, farmers looked for self-sown seedlings and transplanted them (Russell 1955).

The Kola Trade

From time immemorial, the forest and the savanna regions of Africa have been complementary in basic natural products. This complementarity – based on the very different requirements of forest and savanna dwellers – stimulated long-distance exchange between the Volta basin and the Upper Niger (Denham and Clapperton 1826; Lander 1830; Dupuis 1894; Hallet 1965; Bowdich 1966; Arhin 1987). Gold, kola nuts, and salt from the coast constituted the most important items of trade in the Volta Basin of West Africa (Wilks 1962; Fynn 1971).

Evidence from the *Esmeraldo de Situ Orbis* – penned by one of the earliest Portuguese explorers, Duarte Pacheco Pereira – attests to the existence of the West African kola trade as early as the sixteenth century. A materia medica written in 1586 for the Moroccan sultan, Ahmed al-Mansur, mentions that kola was brought from the western Sudan – from “a place called Bitu where there are mines of gold and gold dust” (Wilks 1962: 15).

C. nitida, as an export crop, is reported to have originated in the hinterland of what is now Sierra Leone and Liberia, where it appears to have been commercialized by people who spoke a language of the West Atlantic family of Niger-Congo. People of the Mande branch, who subsequently moved to the forest zone, became

the major producers in the fourteenth century. The kola nut trade was then monopolized by Muslim Mande traders who traveled widely throughout the western Sudan (Goodman et al. 1995). By the fifteenth century, the cultivation of *C. nitida* had spread throughout the Ivory Coast region and the Volta basin. It was also being grown in the forests of Sierra Leone, Liberia, and the Ivory Coast, and in the Volta basin of Ghana.

Coastwise trade in kola led to the creation of two diasporic communities – the Bahun commercial network, linking the lower Cacheu River with the Casamance River and the Gambia River, and a parallel Mandinka network to the east, linking the Upper Geba and Upper Casamance with the Middle and Upper Gambia and Upper Niger (Brooks 1980). Although this kola trade antedated the arrival of the Europeans, when Portuguese traders did reach West Africa, they quickly recognized the commercial importance of kola. With the help of African mariners, they began to participate in the kola trade while systematically concealing that participation from royal regulation and sanction (Brooks 1980).

In the eighteenth and nineteenth centuries, the Gold Coast and Sierra Leone supplied kola to northern and, later, southern Nigeria (*Gold Coast Blue Books* 1884–1925; Miles 1931). During the last half of the nineteenth century, however, demand for Gold Coast kola nuts mounted in Britain (Table III.8.2), France, Germany, and the United States, where the nuts were employed in the production of “neo kola,” cola-based soft drinks, and pharmaceuticals (*Gold Coast Blue Books* 1884–1925).

The overseas kola trade from the Gold Coast to Europe and the United States developed in three phases. It began with the export of two packages of unknown weight to England in 1867 (Dickson 1969) and continued to grow until, in 1884 and 1885, 400 packages of kola, worth more than £2,000, were sent to Europe and the United States (*Gold Coast Blue Books* 1884, 1885; Freeman 1893; Dickson 1969).

Table III.8.2. *Kola exports to England*

1867	2 pkgs
1884	400 pkgs
1885	400 pkgs
1898	1 ton
1899	9 tons
1900	< 9 tons
1901	< 9 tons
1902	< 9 tons
1903	1 ton
1904	1.5 tons
1906	1.6 tons
1907	2 tons
1908	1.5 tons
1910	< 2 tons
1915	3.5 tons

Source: *Gold Coast Blue Books* 1884–1925 and Dickson (1969), pp. 151–3.

The second phase of the southern kola trade, which saw the export of kola to Brazil from the 1890s until the 1920s, took the form of exports directly from Ghana, as well as reexports of Ghanaian kola nuts from Nigeria to Brazil (*Gold Coast Blue Books* 1888–1920; Alex 1890).

The third and final phase of the southern kola trade involved exports to Nigeria in the first decades of the twentieth century. The ports of Accra, Winneba, Saltpond, Apam, and Cape Coast became outlets for the these shipments of kola by sea.

The southern axis of the kola trade was also spurred by the purchase of kola by African and European companies trading along the coast of Ghana and by improvements in communication, such as the construction of roads and railways and the introduction of motor transportation, along with mail boats, steamers, and ferries (Danquah 1928; N.A.G. [Accra] 1959; Dickson 1961). The result was a significant increase in kola trade volume.

From the mid-1920s onward, however, the market demand for Ghanaian kola began to decline. One reason was the beginning of large-scale kola cultivation in Nigeria in the 1920s and 1930s (N.A.G. [Accra] 1959; Cohen 1966; Agiri 1977). But another factor was that cacao, now an international crop, had gradually supplanted kola as the backbone of the Gold Coast economy. It is worthy of note, however, that because of interior African markets, the decline in kola exports to Nigeria did not mean a collapse of kola production in Ghana.

Uses of Kola

In the Past

Initially a luxury item, kola later became an item of common usage. Those who grew it, as well as those who used it, often ascribed therapeutic, dietetic, and pharmacological properties to kola. In addition, according to volume 14 of the *American Druggist* (1885), kola was to some extent employed as a substitute for food, in much the same way as coca has been used by residents of the Andes. Indeed, kola is to the African what tobacco or coffee is to the European or betel is to the southeastern Asian – a stimulant and a psychoactive substance (Simmonds 1891; Kumm 1907; Goodman et al. 1995).

The importance of kola as a drug was first recognized outside Africa in the twelfth century by an Arabian physician, who wrote that it was used in the form of a powder for colic and stomachache and had warming properties. A later Portuguese observer testified to the importance of kola nuts thus: “The Black population would scarcely undertake any enterprise without the aid of kola” (Fluckiger 1884: 524), which, among other things, was supposed to protect against the pangs of thirst. However, the first definite mention of kola as a drug came in the work of Odoard Lopez in 1591, and shortly afterward, Andre Alvares de

Almada, who had visited Guinea in 1566, wrote his *Tratada Breve dos Rios de Guiné do Cabo Verde* (1594), in which he claimed that kola and betel were used in more or less the same way. At the end of the sixteenth century, James Garet, an apothecary and amateur collector of foreign curios, brought the nuts to the attention of the celebrated Flemish botanist-physician Carolus Clusius (Charles de l’Ecluse) (Fluckiger 1884).

As knowledge of kola reached the outside world, the plant itself began to travel, apparently reaching the Caribbean as early as the seventeenth century. This came about following an urgent request, sent through a Guinean trader by an agent for Jamaican slaveholders, for kola seedlings – urgent, because of kola’s well-known property as “a medicinal prophylactic agent or as an ordinary article of food, to avert, as far as practicable, those attacks of constitutional despondency to which . . . Negroes were peculiarly liable” (Attfeld 1865: 456–7).

Many accounts of the properties of kola by early writers were borrowed from travelers and were therefore probably exaggerated or distorted to some degree. Since the 1850s, however, research has been carried out by botanists, chemists, and pharmacists on some of the properties ascribed to the kola nut. For example, A. M. F. J. Palisot-Beauvois (1805) asserted that the Negroes of Oware used kola nuts because of the nuts’ remarkable ability to impart a pleasant taste to all food or water consumed. Subsequent experiments have confirmed this observation, at least for drinking water, which, even when comparatively stale or impure, becomes quite palatable to the consumer after chewing kola. It is possible that the action of the chemicals in kola on the palatal mucosa creates the “illusion” of sweetness, or perhaps this is the result of kola’s high caffeine content.

Another report – this by N. Hudson (1886), a medical inspector in the U.S. Navy – on the results of administering kola paste to a patient suffering from rheumatism, dyspnea, and headache, amply demonstrates some of the pharmaceutical and pharmacological properties of the kola nut. Hudson wrote:

The patient, a lady of 36, had suffered during childhood from rheumatism. Up to a recent period, however, she had been in good health and able to lead a busy, active life. Eighteen months since, she was attacked with severe endocarditis, from which she recovered slowly, and with a damage to the mitral valves. . . . The action of the heart was feeble and irregular, and there was a good deal of dyspnoea, faintness, and fatigue upon even slight exertion. She had always been subject to occasional headaches, and these now became periodic and severe, occurring at first at intervals of four or five weeks, and lasting two days. Latterly they had increased in frequency and intensity, coming on about twice a month, each

attack causing three days of suffering so severe as to fill the patient's mind with constant dread of a recurrence. The discovery of a ratio of urea much below that of normal urine, with the existence of an occasional granular cast, established the conviction that the headaches were uraemic in character (1885-6: 711).

After about three months of administering kola paste in hot milk at a dosage of about 10 grams once or twice daily, Hudson reported:

During the first nine weeks of its use there was no recurrence of headache. Then there was a comparatively mild attack, which may fairly be attributable to a suspension of the use of the kola in consequence of the marked improvement. The general condition has materially improved, the heart's action is more regular, and the attacks of dyspnoea and faintness have nearly disappeared. The most characteristic effect seems to have been an immediate relief of a sense of fatigue, a sense of *bien-être* and cheerfulness to which the patient had been long a stranger. The employment of the kola seemed to be satisfying the appetite, for whenever taken, it appeared to serve as a substitute for the following meal. But the nutritive processes were not impaired: on the contrary, the bodily weight increased from ninety-eight to one hundred and five pounds. No marked change in the character of the urea has been observed. The quantity voided has somewhat increased, but the total urea excreted remains at about ten or twelve grammes daily (1885-6: 711).

As Hudson noted, the results of the administration of kola paste were marked and immediate, producing a definite and positive change in the well-being and comfort of the patient (Hudson 1885-6: 712). Thus, as stated by another observer who was a contemporary of Hudson, kola can become a "sustaining and stimulating adjunct in exhaustive and wasting diseases" (Simmonds 1891: 10).

It was also believed that kola "exercises a favourable influence upon the liver, and that white people, living in those regions [where it grows], who chew a small quantity before meals escape constitutional changes due to affections of that organ" (Heckel and Schlagdenhauffen 1884: 585). Finally, kola was said to be advantageous for sportsmen, athletes, and "brain workers" in reducing tension (Uses of kola nut 1890).

Current Ethnopharmacology in Africa

Kola nuts are used for the treatment of certain infections and conditions, such as guinea worm, migraine, and ulcer (Mr. Ansong, Kumasi Cultural Centre, March 5, 1995, personal communication). In the case of guinea worm, the nut is chewed into a paste and applied to the

affected portion of the body, and in Ghana, such a paste is used to treat a skin ailment commonly called *ananse*. Fresh kola nuts are also chewed as a stimulant to counteract fatigue (Szolnoki 1985; Burrowes 1986). Kola is used in beverages – such as coffee, tea, and cocoa – to cure indigestion and nervous or bilious headache. It should be pointed out that the medicinal action of kola is not the same as that of an analgesic taken for pain. Kola is not used because of an isolated chemical ingredient; it is used as a complex whole.

Along with the nut, other parts of the kola tree are also employed for their medicinal properties by traditional healers and the peoples of rural Africa (Mr. Anin-Agyei, Dwaben State Oil Palm Plantation, February 28, 1995, personal communication; Mr. Forster, Dwaben State Oil Palm Plantation, February 28, 1995, personal communication; Mr. Ansong, Kumasi Cultural Centre, March 5, 1995, personal communication). The bark is used to treat swellings and fresh wounds, and the roots provide excellent chewing sticks for cleaning the teeth. The pod bark is mixed with other ingredients to become a traditional medicine for reducing labor pains during childbirth (Akhigbe 1988). The latter is very important because traditional birth attendants play a dominant role in rural Africa's health delivery system. Most of the people live outside the cities, and the overwhelming majority of these cannot afford the expense of Western-style medicine. Thus, kola-bark preparations for reducing labor pains are highly prized by traditional birth attendants, even though very little research has been done to determine the properties in kola bark that make childbirth more comfortable for African mothers.

Many people use kola because of its very high caffeine content – one nut contains more caffeine than two large cups of American coffee. It is generally believed that chewing small amounts increases mental activity and reduces the need for sleep (Attfield 1865; Ogutuga 1975; Lovejoy 1980). Consequently, kola is widely used by Ghanaian watchmen (who call it the "watchman's friend") in the course of their vigils, as well as by students who work late into the night. Chewing kola also massages the gums and exercises the teeth.

Yet, that kola can be poisonous to animals raises flags of caution. The chemical composition of kola paste makes it a good bait for trapping mice, who often die when they eat it. Similarly, dogs do not survive when they are given *maasa* (a baked flour meal) mixed with kola paste (Mr. Ansong, Kumasi Cultural Centre, January 30 and March 5, 1995, personal communication). There has, however, been no laboratory analysis to explain the action of the kola paste on either mice or dogs.

Social Uses

Kola provides flavoring for beverages in Africa, Europe, North and South America – indeed, in all parts of the world (Freeman 1893; Irvine 1948; Bovill 1958;

Dickson 1969). In the 1870s, kola was mixed with sugar and vanilla as a tonic for invalids and convalescents and was recommended to travelers as an antidote to fatigue and even hunger (Freeman 1893). In 1886, John S. Pemberton, a druggist in Atlanta, Georgia, invented Coca-Cola when he combined coca and kola extracts as a headache and hangover remedy (Louis and Yazijian 1980). Over time, Coca-Cola (minus the narcotic coca) has become the most popular nonalcoholic beverage in the Western world.

The most common social use of kola in several African countries is as a gift of welcome to friends and guests. Doubtless in part, at least, because of the Islamic prohibition against alcohol consumption, kola is widely used by Muslims (Dalziel 1948; Cohen 1966; Dickson 1969; Morgan and Pugh 1969; Church 1975; Agiri 1977; Lovejoy 1980; Anquandah 1982). A Sokoto tradition not only associates kola with the prophet Muhammed but also asserts that he relished it and gave the nuts as gifts to his favorites. His wealthier followers, in turn, gave kola as alms during high festivals (Tremearne 1913, cited in Ferguson 1972; Russell 1955; Lovejoy 1980).

Dietetic Uses

Kola possesses physiological properties that enable those who eat it to undergo prolonged exertion without fatigue or thirst. Kola has also been said to serve as a preventive of dysentery and other intestinal disorders (Fluckiger 1884), and it has been claimed that kola makes people brave in – even eager for – battle. For all these reasons, kola nuts have historically been dispensed to troops on African battlefields. Askia Mahmud is reported to have supplied kola to his troops in the sixteenth century, and at the beginning of the nineteenth century, soldiers of the Asante army often chewed kola for days during their campaigns, when there was frequently not enough food to go around (Kreiger 1954; Quarcoo 1971).

In 1852, the German explorer Heinrich Barth thought that soldiers of the Sokoto army were addicted to kola because it was usually distributed to them in the evenings before campaigns (Goodman et al. 1995). Similarly, James Richardson observed that the Emir of Zinder used kola to incite his soldiers before slave raids. In 1890, the German war office ordered 30 tons of kola for the army after conducting experiments with the plant during autumn maneuvers the previous year (Uses of kola nut 1890). At about the same time, troops of the West African Frontier Force stationed in East Africa were given kola rations to enhance their battle readiness (Anyane 1963).

Toward the end of the nineteenth century, a series of interesting experiments on the possible military importance (in dietetic and pharmacological terms) of kola was conducted by R. H. Firth on British soldiers in India. Ordered by the surgeon-general of Her Majesty's forces, these studies indicated that chewing

kola nuts not only seemed to depress the appetite and keep thirst at bay but also increased the individual's energy level and sense of well-being. In addition, the nuts may have proved efficacious in curing hangovers (Firth 1890).

Nearly a century later, studies carried out on the biochemical effects of kola nut extract administered to rats once again demonstrated the stimulating properties of kola in relatively large doses (Ajarem 1990). This is doubtless explicable in part by the caffeine content; however, many more such efforts are required to show why it is that kola has a long-standing reputation as a laxative (but prevents dysentery and other intestinal disorders) and as a heart stimulant (but can also be a sedative). Also demanding explanation are kola's alleged ability to make foul water palatable and its historical fame as an aphrodisiac, a restorer of potency, and an aid to childbirth (Nzekwu 1961; Quarcoo 1969; Goodman et al. 1995).

Certainly kola seems potent as a drug, but save for the research just mentioned, its potentiality remains to be investigated. Given its many current uses in Africa, as well as its role in the soft-drink industry around the world, it is to be hoped that such research will soon be forthcoming.

Edmund Abaka

Bibliography

- Adamu, Mahdi. 1978. *The Hausa factor in West African history*. Ibadan, Nigeria.
- AERLS (Agricultural Extension and Research Liaison Services). 1982. *Guide to the production of kola*. Extension Guide No. 48 (reprint), Ahmadu Bello University. Zaria, Nigeria.
- Agiri, Babatunde. 1972. Kola production in western Nigeria, 1850–1950. A history of the cultivation of *Cola nitida* in Egba-Owode, Ijebu-Remo, Iwo, and Ota areas. Ph.D. thesis, University of Wisconsin.
1977. The introduction of *nitida* kola into Nigerian agriculture 1880–1920. *African Economic History* 3: 1–14.
- Ajarem, Jamaan S. 1990. Effects of fresh kola-nut extract (*cola nitida*) on the locomotor activities of male mice. *Acta Physiologica et Pharmacologica Bulgaarica* 16: 10–15.
- Akhigbe, F. O. 1988. Kolanut: The characteristics and growth of its literature in Nigeria. *Quarterly Bulletin of the International Association of Agricultural Leadership and Development* 33: 47–52.
- Alex, Geo. 1890. To the Marquis of Salisbury, K.G., September 6, 1890. *Royal Botanical Garden, Kew, Bulletin of Miscellaneous Information* 4: 253–60.
- Anquandah, James. 1982. *Rediscovering Ghana's past*. Accra and London.
- Anthony, K. R. M., Bruce F. Johnston, William O. Jones, and Victor C. Uchendu. 1979. *Agricultural change in tropical Africa*. Ithaca, N.Y., and London.
- Anyane, S. La. 1963. *Ghana agriculture. Its economic development from early times to the middle of the twentieth century*. London.

- Archibald, J. F. 1955. *The propagation of cacao by cuttings*. West African Cocoa Research Institute, Technical Bulletin No. 3. Nigeria.
- Are, L. 1965. Performance of cacao from roughly handled rootstocks grown under different conditions. *Nigerian Agriculture Journal* 2: 61-5.
- Arhin, Kwame. 1977. *West African traders in the nineteenth and twentieth centuries*. London.
1987. Savannah contribution to the Asante political economy. In *The golden stool: Studies in the Asante center and the periphery*, ed. Enid Schildkrout, 51-9. New York.
- Ashiru, G. A. 1969. Dioxide, light and heat on seed germination and seedling growth of kola (*Cola nitida* [Ventenant] Schott & Endlicher). *American Society for Horticultural Science Journal* 94: 429-32.
1975. The vegetative propagation of tropical tree crops with emphasis on cacao (*Theobroma cacao* L.) and kola (*Cola nitida*, [Vent.] [Schott & Endlicher]). *Acta Horticulturae* 49: 67-73.
- Ashiru, G. A., and T. A. Quarcoo. 1971. Vegetative propagation of kola (*Cola nitida* [Vent.] [Schott & Endlicher]). *Tropical Agriculture* (Trinidad) 48: 85-92.
- Attfield, John. 1865. On the food value of the kola nut - a new source of theine. *Pharmaceutical Journal* (January): 457-60.
- Berberich, Charles William. 1974. A locational analysis of the trade routes of the Asante north east frontier network in the 19th century. Ph.D. thesis, Northwestern University.
- Boahen, Adu. 1964. *Britain, the Sahara and the western Sudan*. Oxford.
- Boaten, Nan Akwasi Abayie I. 1990. Asante: The perception and utilization of the environment before the twentieth century. *Research Review* 6: 19-28.
- Bodard, M. 1954. Polyembryo in *Cola nitida*. In *Proceedings of the 8th International Congress of Botany*, 259-61. Paris.
- Bovill, E. W. 1958. *The golden trade of the Moors*. London.
- Bowdich, T. E. 1966. *Mission from Cape Coast to Ashantee*. Third edition. London.
- Brooks, George. 1980. *Kola trade and state building: Upper Guinea coast and Senegambia 15th to 17th centuries*. Boston University Working Papers, No. 38. Boston, Mass.
- Brown, D. A. Li. 1970. A review of germination of kola seed (*Cola nitida* [Vent.] Schott & Endlicher). *Ghana Journal of Agricultural Science* 3: 179-86.
- Burrowes, C. P. 1986. Economic activities of pre-Liberian societies: Production for local use and exchange. *Liberia Forum* 2: 25-44.
- Chevalier, Auguste, and Emile Perrot. 1911. *Les kolatiers et les noix de kola*. Paris.
- Church, R. J. Harrison. 1975. *West Africa: A study of the environment and man's use of it*. London.
- Cohen, Abner. 1966. The politics of the kola trade. *Africa* 36: 18-35.
- Curtin, Philip. 1975. *Economic change in precolonial Africa: Senegambia in the era of the slave trade*. Madison, Wis.
- Daaku, K. Y. 1970. *Trade and politics on the Gold Coast, 1600-1720*. Oxford.
- Dalziel, J. M. 1948. *The useful plants of west tropical Africa*. London.
- Danquah, J. B. 1928. *Akyem Abuakwa handbook*. London.
- Denham, Dixon and Hugh Clapperton. 1826. *Narrative of travels and discoveries in Northern Central Africa, in the years 1822, 1823 and 1824*. London.
- Dickson, K. B. 1961. The development of road transport in southern Ghana and Ashanti since about 1850. *Transactions of the Historical Society of Ghana*, 33-42.
1969. *A historical geography of Ghana*. Cambridge.
- Dupuis, Joseph. 1894. *Journal of a resident in Ashantee*. London.
- Fage, J. D. 1969. *A history of West Africa*. London.
- Ferguson, Phyllis. 1972. Islamization of Dagbon: A study of the Alfanema of Yendi. Ph.D. dissertation, Cambridge University.
- Firth, R. H. 1890. The dietetic and therapeutic value of the kola nut. *The Practitioner* 43 (Supplement): 27-35.
- Fluckiger, F. 1884. A contribution to the history of kola. *The New Idea* (Detroit) September: 224.
- Freeman, Austin. 1893. A journey to Bontuku in the interior of West Africa. *Royal Geographical Society Supplementary Papers* 3: 117-46.
- Freeman, Richard Austin. 1967. *Travels and life in Asbanti and Jaman*. London.
- Fynn, J. K. 1971. *Asante and its neighbours, 1700-1807*. Evanston, Ill.
- Gnanaratum, J. K. 1972. Rooting of cacao cuttings, a new technique. *World Crops* 16: 61-3.
- Gold Coast Blue Books*. 1884-1925.
- Goodman, J., P. E. Lovejoy, and A. Sheratt, eds. 1995. *Consuming habits. Drugs in history and anthropology*. London and New York.
- Goody, Jack. 1967. The over-kingdom of Gonja. In *West African kingdoms in the nineteenth century*, ed. Daryll Forde and P. M. Kaberry, 179-205. London.
- Hallet, Robin, ed. 1965. *The Niger journal of Richard and John Lander*. London.
- Heckel, E., and F. Schlagdenhauffen. 1884. Some African kolas in their botanical chemical and therapeutic aspects. *The Pharmaceutical Journal and Transactions* (January 26): 584-6.
- Hudson, N. 1885-6. Kola. *Medical Times and Register* 16: 711-12.
- Hunter, Guy. 1973. Agricultural administration and institutions. *Food Research Institute Studies* 12: 233-51.
- Ibikunle, B. A. O. 1972. The propagation and growth of nursery stock of *Cola nitida* (Vent.) (Schott & Endlicher). Ph.D. thesis, University of Ibadan, Nigeria.
1975. The germination of *Cola acuminata*. *Acta Horticulturae* 49: 75-83.
- Irvine, F. R. 1948. The indigenous food plants of West Africa. *Journal of the New York Botanical Garden* 49: 225-36.
1969. *West African crops*. London.
- Jones, Norman. 1968. Plant breeding: Problems with forest tree species in West Africa. *Ghana Journal of Agricultural Science* 48: 27.
- Jones, Stephen Hugh. 1995. Coca, beer, cigars, and yagé: Meals and anti-meals in an Amerindian community. In *Consuming habits. Drugs in history and anthropology*, ed. Jordan Goodman, P. E. Lovejoy, and A. Sheratt, 47-66. London and New York.
- Karikari, S. K. 1973. The effect of maturity and storage on the germination of kola nut, (*Cola nitida* [Ventenant] Schott and Endlicher). *Ghana Journal of Agriculture Science* 49: 225-36.
- Kennedy, John G. 1987. *The flower of paradise. The institutionalised use of the drug qat in North Yemen*. Dordrecht, the Netherlands.
- Kola. 1957. *The Ghana Farmer* 1: 222.
- Kraemer, Henry. 1910. *A text-book of botany and pharmacognosy*. Philadelphia, Pa.

- Kreiger, Kurt. 1954. Kola Karawanen. *Mitteilungen des Institutes für Orientforschung, Deutsche Akademie der Wissenschaften zu Berlin* 2: 289-322.
- Kumm, H. Kail. 1907. *The Sudan*. London.
- Lander, Richard. 1830. *Records of Capt. Clapperton's last expedition to Africa. With the subsequent adventures of the author*. London.
- Louis, J. C., and Harvey C. Yazizjian. 1980. *The cola wars*. New York.
- Lovejoy, Paul E. 1977-8. Kola in West Africa. *Cahiers d'études africaines* 20: 97-134.
1980. *Caravans of kola. The Hausa kola trade 1700-1900*. London and Zaria, Nigeria.
1995. Kola nuts: The "coffee" of the central Sudan. In *Consuming habits. Drugs in history and anthropology*, ed. J. Goodman, P. E. Lovejoy, and A. Sheratt, 103-25. London and New York.
- Miles, A. C. 1931. *Report of a cola survey of the eastern Asbanti areas and a general review of the Gold Coast industry, 20 March 1931*. London.
- Morgan, W. B., and J. C. Pugh. 1969. *West Africa*. London.
- N.A.G. (National Archives of Ghana) (Accra). 1959. *Gold Coast Colony, report of the Department of Agriculture for the year 1930-31*. CSO 8/1/59.
- N.A.G. (National Archives of Ghana) (Kumasi). 1905. D.1346, Botanic Gardens, Aburi-Gold Coast, 30/11/1905.
- Nzekwu, O. 1961. Kolanut. *Nigeria Magazine* 71: 298-305.
- Ogutuga, D. B. A. 1975. Chemical composition and potential commercial uses of kola nut, *Cola nitida* Vent. (Schott & Endlicher). *Ghana Journal of Agricultural Science* 8: 121-5.
- Oladokun, M. A. O., and N. O. Adepipe. 1989. Polyembryony and multiple shoot production in *cola acuminata* (P. Beauv) Schott and Endlicher. *Nigeria Journal of Botany* 2: 135-42.
- Oliver, Daniel. 1868. *Flora of tropical Africa*, Vol. 1. Asford, Kent.
- Osim, E. E., and P. M. Udia. 1993. Effects of consuming a kola nut (*Cola nitida*) diet on mean arterial pressure in rats. *International Journal of Pharmacology* 31: 192-7.
- Ossei. 1963. Cola - some aspects of its cultivation and economics in Ghana. *Ghana Farmer* 7: 96-7.
- Palisot-Beauvois, A. M. F. J. 1805. *Flore d'Oware et de Benin en Afrique*. Paris.
- Pratt, Robertson, and Heber W. Youngken, Jr. 1956. *Pharmacognosy: The study of natural drug substances and certain allied products*. Philadelphia, Pa.
- Pyke, E. E. 1934. A note on the vegetative propagation of kola (*Cola acuminata*) by softwood cuttings. *Tropical Agriculture* (Trinidad) 11: 4.
- Quarcoo, T. A. 1969. Development of kola. Its future in Nigeria. *Proceedings of the Agricultural Society of Nigeria* 6: 19-22.
1971. Processing of kola in Nigeria. In *Progress in tree crop research in Nigeria (coca, kola and coffee)*. Cocoa Research Institute. Ibadan, Nigeria.
- Ramstad, Egil. 1959. *Modern pharmacognosy*. New York.
- Royal Botanical Garden, Kew. 1890. Cola Nut. *Bulletin of Miscellaneous Information* 4: 253-60.
1906. Cola. *Bulletin of Miscellaneous Information* 4: 89-91.
- Russell, T. A. 1955. The kola of Nigeria and the Cameroons. *Tropical Agriculture* 32: 210-40.
- Simmonds, P. L. 1891. The kola nut of Africa. *American Journal of Pharmacy* 62: 595-7.
- Staff of the Division of Agriculture. 1961. Crops other than cocoa and the diseases and pests which affect them. In *Agriculture and land use in Ghana*, ed. J. B. Wills, 360-1. London and New York.
- Szolnoki, T. W. 1985. *Fruit and fruit trees of the Gambia*. Hamburg.
- The kola nut in western Soudan. 1890. *Pharmaceutical Record* (May): 181.
- Toxopeus, H., and G. E. Okoloko. 1967. Cacao marcotting in Nigeria. I. Use, technique and monthly rooting success. *Nigerian Agricultural Journal* 4: 45-8.
- Uses of kola nut. 1890. *Pharmaceutical Record* 10 (June): 226.
- Wilks, Ivor. 1962. A medieval trade route from the Niger to the Gulf of Guinea. *Journal of African History* 3: 337-41.

III.9 Milk and Dairy Products

Purity and Danger

Milk occupies a curiously ambiguous place in the history and culture of food. It has been pointed to as an archetypal, almost elementally nourishing food, supremely healthful, reflecting the nurturing relationship of mother and infant. In recent times, its whiteness has come to stand as a symbol of natural goodness and purity. But milk also conceals danger. Its nutritional largesse is equally appealing to hosts of putrefying bacteria, and unless milk is consumed almost immediately, it rapidly deteriorates into a decidedly unwholesome mass. Even in the apparently safe period between lactation and curdling, pathogenic organisms may lurk and multiply with potentially more devastating consequences for a new infant than the more immediately apparent problems of an obviously bad food.

The very processes of corruption, however, also provided the ways by which milk became a more widespread and acceptable food. Some contaminating organisms transform milk toward simple forms of butter, cheese, or yoghurt, and it is in these forms, not as a beverage, that milk has been consumed throughout the greater part of the history of human eating. As a highly ephemeral food then, unless milk is transmitted directly between provider (whether human or animal) and consumer, it is fraught with danger. Preservation has thus been the overriding factor in milk's development as an important food for humans.

Initially, preservation was achieved through manufacture into butter or cheese. Later, briefly, fresh milk was kept safe only by cleanliness of production and speed of transport; in the twentieth century, however, milk has been preserved primarily by means of heat treatment, particularly pasteurization. This preservation of milk, particularly on an industrial scale since the late nineteenth century, highlights another contradictory tension in the nature of its consumption. Milk

production is a quintessentially female process, and the resonances of the mothering bond imparted a crucially feminine nature to the whole area of dairying in preindustrial times. Even though milk could only become an important foodstuff by transforming the female milk into a different, harder manufactured solidity, commercial dairies and domestic output remained spheres of women's activity. The femininity of dairy production, however, could not withstand industrial dairying methods, and from the end of the nineteenth century, men began to take over dairying as it became concentrated in bigger, more technically sophisticated industrial plants.

There is a further fundamental dichotomy in the culture and history of milk (one almost unique among foodstuffs) in the very ability of people to digest it. Some people do not have the enzyme – lactase – required to digest lactose, the principal sugar in milk, and the pain and discomfort arising from the inability to absorb milk sugars are reflected in cultural traditions that render even the thought of imbibing milk repulsive. Lactase deficiency affects most of the world's peoples, many of whom regard milk drinking as revolting. Thus, milk has been geographically limited, even though it is universally available. It has been consumed in parts of Asia and Africa, but its consumption has been most significant in Europe and in areas of European colonization, such as Australasia and North America.

These tensions and contradictions in the very nature of milk – a pristine, white, and nutritious beverage, which also harbors and conceals corruption; a whole and complete food which, in principle, demands no preparation but which, in practice for most of its history, has required transformation into something else, and which vast numbers of people literally cannot stomach – are reflected in the history of milk production and consumption. In preindustrial societies, dairy produce had a part to play in peoples' diets, although it was probably not particularly important in terms of overall nutrition. Yet, milk had a cultural resonance, beyond its dietary value, that saw it enshrined in several religious traditions as a signifier of promise and plenty, or as a symbolic link between spiritual and earthly nourishment. From the early eighteenth-century commercialization of agriculture, dairying took on greater economic importance, particularly in provisioning urban centers.

In the nineteenth century, increasing urbanization and the development of rapid transportation and communication systems, particularly railways, began for the first time to create a market for liquid milk on a large scale. Dairying and the production of liquid milk acquired nationalistic significance around the beginning of the twentieth century as a result of growing concern in Western countries about the health of infants. Yet although milk had come to be vitally important, the uncertain hygiene associated with it highlighted the necessity of clean production

and supply. The original dichotomies regulating the consumption of milk remained, however, because the dissemination of liquid milk required enormous mechanical intervention, and its natural purity could be maintained only by industrial processing.

Throughout the twentieth century, as milk eventually became a readily available, clean, and safe commodity, its consumption as a beverage reached a cultural epitome: Its whiteness evoked an image of goodness – a signifier of health and hygiene. Its ascendancy was short-lived, however; in the overfed northern hemisphere, the same nutritional value that had made it so important in the previous decades consigned it to the ranks of dietary sins. At the same time, inappropriate developing-world feeding schemes (based on dried milk) and the recognition of lactase deficiency have undermined the notion of universal goodness and revealed the dangers of a nutritional imperialism.

Milk in Preindustrial Societies

To drink milk is an inherently natural impulse; an infant placed at the breast knows instinctively what to do. As suitable animals (such as mares, asses, ewes, goats, and various kinds of cows) were domesticated, it appears that drinking animal milk became an acceptable practice. Presumably, folk taxonomic associations between humans and other mammals would have indicated that milk was an animal secretion that could be consumed as food, and there are numerous records and legends of infants suckling directly from animals, as well as from a range of artificial devices (Fildes 1986).

As with most other early foods, the origins of dairy products are unclear. Certain natural fermentations give rise to yoghurt or soft, cheeselike substances that are sufficiently different from putrid milk to have encouraged their sampling. There are stories that cheese originated among Near Eastern nomads who may have stored milk in the stomach bags of cows, a common form of container, in which the natural rennet would have given rise to a sort of cheese. Similarly, one might imagine a horseman setting out with a bag of milk, only to find, when he came to drink it, that the agitation of riding had curdled it into a not unpleasant mixture of curds and buttermilk (Tannahill 1973). Whatever the origins, dairy products have had a place in human diets from very early times.

Equally early on in human history, however, a fundamental split occurred in the history and culture of dairy consumption because, as already noted, drinking more than very small quantities of milk causes the majority of the world's adult population to suffer digestive problems. It has been estimated that although more than 96 percent of northern European peoples are able to digest milk, some 50 to 75 percent of Africans, Indians, Near Eastern Asians, eastern Euro-

peas, and virtually all Asian and Native American peoples cannot digest it. Their bodies stop producing lactase – the enzyme that breaks down the milk sugar, lactose – soon after weaning (Tannahill 1973). It has been suggested that an adult ability to break down lactose spread as people moved northward into colder climates. Covering themselves with more clothes, these people experienced shortages of vitamin D previously derived from the action of sunlight. Vitamin D production in the skin, however, is enhanced with a higher intake of calcium, of which milk is a particularly rich source. There would, thus, have been a selective advantage for those people who retained the capacity to digest milk, and over time, the proportion of the population in northern climates with the ability to digest lactose would have increased (Harris 1986).

Such a starkly biological account, however, does not explain the powerful rejection of milk among many Asian peoples and needs to be supplemented with cultural factors. One suggestion is that Chinese agricultural practices worked to exclude milk from the culture; because planting took place year-round, primarily with human labor in areas of high population density, there were few draft animals that could have provided milk. The main flesh animal was the pig, which was impossible to milk; thus, after weaning, Chinese toddlers were not exposed to milk from other sources. By contrast, in subcontinental Asia, draft animals were more prevalent – indeed, were necessary to prepare land for the shorter planting season dictated by a monsoon climate. Because of greater availability of milk, cultural aversion to it did not develop, and dairy products remained an important feature of the diet (Harris 1986). An alternative hypothesis, and one, apparently, traditionally held by Chinese people, is that the aversion to milk arose from the desire to distinguish themselves from the nomads on the northern borders of China, who drank fermented mare's milk (Chang 1977). It does seem that milk products had become a feature of the diet of the northern Chinese aristocracy between the Han and Sung periods, but it appears that, after the ensuing period of Mongolian rule, milk definitely acquired a barbarian reputation (Chang 1977).

In most of the preindustrial world, however, milk (or, more usually, dairy products) had a place in the diet of all those who had domesticated animals and could absorb lactose. Among nomadic pastoralists of the Near East, sheep and goats provided the main source of milk and cheese. On the great open grasslands of central Asia, and extending into eastern Europe, mare's milk, often fermented into the alcoholic liquor *kumiss*, was consumed in large quantities (Tannahill 1973). On subcontinental Asia, the buffalo was the principal source of milk to be made primarily into ghee – a reduced butter in which most of the moisture is removed by heating – for cooking or ceremonial purposes (Mahias 1988). Pastoralists of Africa,

with vast herds of cattle, used a considerable amount of milk in their diets (Pyke 1968). Milk and cheese are also mentioned as foods, medicines, and beauty aids in records of the ancient civilizations of Egypt, Greece, and Rome (Warner 1976).

In Europe, dairy products were a notable part of the peasants' diet wherever they were available. Across the continent, the great majority of people lived on the verge of starvation, particularly during the late winter months and the periodic harvest failures. The basic diet of cereal-based porridge or soup was rarely supplemented with actual meat, and the more prevalent dairy products (or "white meats," as they were called) were welcome sources of animal fat and protein, whether as cheese, butter, beverage, or additions to soup (Mennell 1985). In the early months of the year, however, cheese or butter kept throughout the winter was mostly used up, and cattle not slaughtered the previous autumn struggled to survive until the new grass appeared (Smith and Christian 1984). Thus, though an important source of animal fat and proteins, dairy products (like meat) cannot have been much more than a flavoring for the basically cereal diet of the peasants.

Among the nobility in the late medieval period there was a growing disdain for dairy produce. With upper-class access to often vast quantities of meat, the "white meats" of the peasantry were increasingly scorned, probably especially by the aspiring merchants of the towns. Nonetheless, dairy products did not disappear from the tables of the well-to-do, and in the late sixteenth and early seventeenth centuries, butter became more important in noble larders, although used primarily for cooking and making the elaborate sauces of the new haute cuisine then developing in France and Italy (Mennell 1985).

Perhaps because milk was recognized as a particularly rich and nourishing food, as well as something of a luxury to have in more than small quantities, but also (and one suspects primarily) because of the symbolic importance of the nurturing bond between mother and child, milk has achieved some prominence in myths and religious systems. In the Old Testament, the promised land was one of "milk and honey." The image of a mother goddess suckling her child is a common representation of Earth bringing forth sustenance. The Egyptian mother goddess, Isis, suckled her divine son Horus, while the Greek god Zeus was nurtured by Amalthea (variously depicted as a mortal woman or as a nanny goat). The symbolism of the mother with her infant god passed into Christian representations of the Madonna and Child, which perpetuated the divine linkages between the spiritual and earthly worlds, mediated by the physical nurturing of milk (Warner 1976). Milk and dairy products have a vital role in sacrificial and purifying rituals in Indian religious myths, especially as the life-giving force of the fire god Agni (Mahias 1988).

The place of dairy products in the diet of peoples in nonindustrial societies has remained virtually unchanged over time; they can be an important source of animal protein but remain a minor component of largely cereal or vegetable diets. In northern Europe, however, and particularly in England, the role of milk and dairy products began to change with the wider emergence of commercial agriculture, urbanization, and proto-industrialization, beginning about the end of the sixteenth century and continuing throughout the seventeenth century.

In seventeenth-century England, demand for dairy products increased with the emergence of rural industry and the location of an increasingly higher proportion of the population in towns (Drummond and Wilbraham 1939). Cheese was of growing importance in the diets of urban laborers, particularly in London. It was cheap, nutritious, and also convenient; it could be stored for reasonable lengths of time and could be carried to places of work and eaten easily there. Large quantities of cheese were also required by the navy and by new, large institutions, such as hospitals and workhouses (Fussell 1926-9; Fussell and Goodman 1934-7). This demand was largely met by the commercialized agriculture that had been developing on the larger estates carved out after the Reformation. The numbers of milch cows multiplied, and dairies became necessary and integral parts of English country houses. Dairying was practiced by anyone who had access to a cow, which replaced the sheep as the primary milk provider (Tannahill 1973), and, for the respectable poor in rural areas, could provide an important source of income. The cheese trade also relied on a reasonably efficient transportation system, and London was served by sea as well as by road (Fussell 1966).

Throughout the early phases of expanding commercialization, dairying remained a female domain characterized by an arcane knowledge (Valenze 1991). Although from southern Africa to northern Europe the notion persisted that a woman who handled milk during menstruation might curdle it, the mystery of dairying lay in the special competence of dairymaids (Fussell 1966; Pyke 1968). The dairy in a country house was invariably attached to the kitchen, supervised by the farmer's wife (or the female head of the household in larger concerns), and the production of cheese and butter was a female operation. From the seventeenth century, books of household management included dairying as a routine aspect of the mistress's responsibilities. Hygiene in the dairy was constantly stressed, with detailed instructions provided as to the proper construction and maintenance of the dairy and the duties of the women working in it (Fussell 1966).

By the end of the eighteenth century, wherever dairying continued to be carried out in a preindustrial, subsistence agricultural context, milk products retained their customary position as a minor adjunct

to the diet, varying only by degrees among places where dairying was more or less pronounced. But in the industrializing and urbanizing world, large-scale commercial agriculture was beginning to alter the nature of production and consumption of dairy produce and bring about a crucial transformation in the historical culture of milk.

Milk in an Urbanizing World

The tripartite revolution of industrialization, urbanization, and the commercialization and increasing productivity of agriculture had dramatic consequences for food production and consumption. Enough food was produced to fuel enormous population growth, with increasing proportions of that population ceasing to be agriculturally productive. Urban dwellers, as net food consumers, depended on food being brought in from producing areas. Throughout the nineteenth century, this took place not only on a national but an international scale, until most parts of the world became integrated into a global network of food trade, principally geared toward feeding Europe. Thus, surplus dairy producers sent huge quantities of cheese and butter, increasingly manufactured in factories, to northern Europe.

At the same time, there was a growing demand for liquid milk in urban areas, and nearby dairy producers began to concentrate on its supply, partially to offset the competition in manufactured products. Yet, accompanying this expansion of production was an increasing consumer concern about the quality of the milk supply, particularly toward the end of the nineteenth century, when milk was implicated in questions of infant mortality. In the elaboration of a range of measures to control problems of milk adulteration and unhygienic procedures, dairying was completely transformed into a highly mechanized manufacturing and distribution activity, which steadily undermined its traditionally feminine nature. Such a pattern recurred throughout the nineteenth-century developing world - commercial, industrial dairying geared toward a manufactured dairy market, paralleled by the development of liquid milk production in urban areas, which, in turn, gave rise to concern about milk quality.

During the eighteenth century, northern European agriculture became much more productive, with novel crop rotation techniques, new machinery, better land management, and the more intensive methods achieved with enclosure. Dairying benefited from greater attention to fodder crops, which allowed cows to be fed more adequately throughout the winter. Yields still fell during winter months, but more cows survived, and yields quickly recovered with the spring grass (Fussell 1966). Thus, the potential arose for year-round dairying.

By the second half of the eighteenth century, increasing food production was supplying not only a

steadily growing rural population but also, in Britain, an explosive increase in urban populations. During the first half of the nineteenth century, the sources of Britain's wealth gradually shifted from agriculture and the landed estates toward manufacturing and industry. This was symbolized by the repeal in 1846 of the Corn Laws, which saw landed and agricultural interests supplanted by the demands of industry for free trade. Yet, as Britain ceased to be agriculturally self-sufficient, free traders had to cater not only to the requirements of industry but also to the need for imported foodstuffs to feed urban industrial workers (Burnett 1989).

Across the globe, producers of agricultural surplus geared up to meet this need (Offer 1989). Australia and New Zealand sent butter and meat to British markets; Ireland and Denmark developed dairying to meet British demand. Danish farmers, recognizing the likely requirements of their rapidly urbanizing neighbor across the North Sea, organized in cooperatives to develop highly efficient methods for producing butter and bacon. In the process, Denmark's national economy became modernized and export driven, yet still based on agriculture (Murray 1977; Keillor 1993).

Ireland had been characterized by very high dairy production and consumption patterns from medieval times, although by the eighteenth century, the continued reliance on dairy foods may be seen as a mark of poverty. As was the case in Denmark, nineteenth-century Irish dairying became more commercialized to supply Ireland's urbanized neighbor, though at the expense of an important source of animal food for its own consumption (O'Grada 1977; Cullen 1992).

In the United States, between the end of the eighteenth century and the middle of the nineteenth, westward expansion had brought new, high-yielding agricultural lands into production. By midcentury, cheese was being shipped down the Erie Canal from upstate New York. But across the vast open spaces of North America, rail transport was the key to the development of dairying. Also crucial was the hand cream separator (Cochrane 1979), which allowed even small farmers to cream off the butterfat that railroads could then take to regional factories to be made into cheese and butter. Industrial methods for manufacturing these products spread across the northeastern and north central states (Lampard 1963; Cochrane 1979). By the second half of the nineteenth century, production and transportation were such that cheese manufactured in the United States or Canada, and butter made in New Zealand, could reach British markets at lower prices than those required by British farmers (Burnett 1989). In addition, an outbreak of rinderpest in England during the 1860s wiped out most of the urban cows, which helped North American dairy products gain a prominent place in British markets (Fussell 1966).

As in the United States, railways in Great Britain were fundamental in enhancing the importance of liq-

uid milk as a beverage. From the 1840s, liquid milk could be brought from the countryside to the towns and sold before it became sour. The perishability of milk had always restricted its scope as a drink. But speed, together with the development of coolers and refrigerated railway cars, increased its viability, and in all areas within reach of an urban center, dairying was increasingly concerned with liquid milk supply. By the second half of the nineteenth century, milk was being carried to London from as far away as Derbyshire, and Lancashire and Cheshire emerged as major dairy counties to supply the conurbations of Liverpool, Manchester, and Stoke (Whetham 1964; Taylor 1974, 1976, 1987). Effectively, the whole of Britain could now be regarded as an urban area, which had an enormous demand for milk.

In North America, the manufacturing of dairy products was concentrated in the lake states of Wisconsin, Minnesota, and Illinois, whereas liquid milk to supply major urban areas was produced in the northeastern and Atlantic seaboard states (Lampard 1963; Cochrane 1979). Thus, better communications and large-scale manufacturing prompted a functional division in dairying. Although manufactured goods, such as cheese, could be carried long distances to urban markets more cheaply than those of a small, local producer, dairying in urban areas enjoyed a protected market for liquid milk that was not susceptible to more distant or foreign competition.

In England, toward the end of the nineteenth century, liquid milk came to be seen almost as an agricultural panacea, and not just for hard-pressed dairy farmers in urban areas. During the great depression of English farming from the mid-1870s to the 1890s, when the full economic impact of imported food began to be felt, farmers increasingly turned to the naturally protected production of liquid milk. More acres reverted to pasture, which was condemned by commentators bemoaning the apparent decline of cereal-based farming, but liquid milk production provided farmers with a welcome respite from foreign competition and helped alleviate the agricultural slump, furnishing a regular cash income throughout the year, helping to clear debts, and bringing some semblance of profitability (Taylor 1974, 1987).

Although few other nations relied so heavily on imports as Britain, by the end of the century, farmers in many countries were feeling the effects of competition in the international food market. In the United States, grain farmers of the mid-nineteenth century found themselves threatened by even higher-yielding lands opening up in the west and sought to diversify, notably into dairying. Delegates were sent from the north central states to study Danish methods, and many Danes forged new careers in the United States, ultimately to the consternation of the Danish government, which feared rivalry for the British market (Keillor 1993). State agricultural colleges and the new experiment stations also sought to improve upon the

(popularly perceived) low standards of American dairying. Scientific "book farming," popular with progressives, found an ideal outlet in dairying, with rational and efficient farming methods based on calculated feeding plans, milk recording schemes, and analyses of butterfat content (Johnson 1971; Peagram 1991). Dairying was becoming self-consciously modern.

Some of the grain states, where there has been an overreliance on single crops, saw dairying as a useful means of diversifying. The North Dakota agricultural experiment station, for example, tried in the early twentieth century to offset the state's reliance on spring wheat by promoting dairying. Although for a while North Dakota did become a notable butter producer, it seems that cereal farmers did not adjust happily to the rather different demands of dairy farming (Danbom 1989).

Ultimately, the general pattern of dairying common to urbanizing countries was established in the United States. Liquid milk was the mainstay of farmers with access to urban markets, and the manufacture of other dairy products was concentrated in the north central states (Michigan, Minnesota, and Wisconsin) on the principal rail routes to the East (Haystead and File 1955; Lampard 1963; Cochrane 1979).

The growing industrialization of dairying and the manufacture of dairy products steadily eroded the femininity of dairying. Butter production, especially, had remained an essentially female activity throughout most of the nineteenth century, occupying an important role not only in the domestic economy but also in the wider social relations and status of women in farming communities. Household manufacture of milk products, however, was increasingly replaced by the transportation of milk to railways and factories and by industrial production (all male spheres of activity), and a marked decline occurred in the number of women involved in dairying. In a typically paradoxical process, the more natural, feminine state of liquid milk gained greater prominence, but only through the intervention of mechanical artifacts operated by men. As dairying left the household, an important element of rural female employment, skill, and authority went with it (Cohen 1984; Osternd 1988; Nurnally 1989; Bourke 1990; Valenze 1991).

Also embattled were milk consumers, increasingly concentrated in urban centers, subject to the vagaries of transport systems for their food supply, and suffering the appalling conditions of massive urban expansion. Many people living in towns simply did not have enough to eat; what was available was of indifferent to abominable quality and, frequently, heavily adulterated. Milk and dairy products were part of the diet, but their supply was highly uncertain. The principal sources were urban milk shops, with cowsheds that provoked bitter condemnation from health reformers. Such places were notorious for squalid conditions, with the cows fed on slops and refuse (frequently the

spent grain from distilleries), and disease rife among them (Okun 1986; Burnett 1989). Moreover, milk, whether from urban shops or from roundsmen coming in from the country, was routinely adulterated in the mid-nineteenth century (Atkins 1992). Arthur Hill Hassall, investigating for the *Lancet* in the early 1850s, found that milk was diluted with from 10 to 50 percent water, with that water often from polluted wells and springs (Burnett 1989).

Throughout the second half of the nineteenth century, urban diets in Britain improved noticeably, both in quantity and quality. This was primarily a result of the rise in real wages as general economic prosperity began to percolate down the social strata, but it was also because of an increasing availability of cheap imported food. Food had always been the principal item of expenditure for urban workers, and so any extra real income was invariably spent on food first, both to extend the variety of the diet and to obtain greater quantities. Thus, from the 1890s, more meat, eggs, dairy produce, and fresh vegetables appeared on the tables of the urban working classes (Oddy and Miller 1976, 1985).

The quality of food was also addressed in a more effective fashion. During decades of revelations by such individuals as Frederick Accum about the extent of food adulteration, a series of food purity laws were introduced in Britain. These were initially rather ineffective, but the 1875 Sale of Food and Drugs Act, and the Public Health Act of the same year, began to bring about real improvements in the basic quality of foods (Burnett 1989). This story was a familiar one in most major urban nations during the second half of the nineteenth century. In the United States, for example, reformers (also stimulated by Accum) launched investigations and turned up their own evidence of food adulteration. Effective legislation was introduced by cities and states by the end of the century, culminating in the 1906 Pure Food and Drugs Act (Okun 1986).

Urban populations probably benefited from the volumes of liquid milk brought in by the railways, but fresh milk still became less than fresh very quickly, and although gross adulterants were being eliminated, the problem of the keeping properties of milk remained. Dairy-producing countries had sought means of preserving milk throughout the nineteenth century. During the 1850s, developments in condensing techniques, especially by the American Gail Borden, brought about the formation, in 1865, of the Anglo-Swiss Condensed Milk Company, which soon had factories across Europe. Tinned milk consumption rose rapidly after 1870, and in the early twentieth century, several methods for making dried milk encouraged the emergence of a powdered-milk industry. In New Zealand, a dairy exporter developed the new Just-Hatmaker process of drying milk and this, in turn, gave birth to the giant corporation Glaxo (Davenport-Hines and Slinn 1992). In England, the pharmaceutical

company Allen and Hanbury's used a method of oven-drying evaporated milk (Tweedale 1990). Condensed and powdered milk were both popular, as they kept longer than fresh milk; moreover, tinned milk could be diluted more or less heavily according to the vicissitudes of the family economy.

By the beginning of the twentieth century, adult diets had improved (although women remained less fed well into the new century), and the worst excesses of urban squalor, poverty, and deprivation were being addressed. These improvements of diet, sanitation, and housing were reflected in the general falling of mortality rates (McKeown 1969), but infant mortality rates remained stubbornly high. The problem was particularly noticeable in France (which also had a declining birth rate) after 1870 and in Britain during the 1890s. Yet concern for high infant mortality was also expressed in Canada, the United States, Australia, New Zealand, and the Netherlands. The underlying issue was the same: With an increasingly tense international situation brought on by the formation of rival power blocks in Europe and the imperial maneuverings by the United States, greater significance was accorded to the health of a nation's people. Populations began to be seen as national assets, and both their sizes and quality were viewed as of national political, economic, and military importance. Attention was focused on the health of urban populations and, particularly, the health of infants who would be the soldiers, workers, and mothers of the future.

Beginning in France in the 1890s, governments, charities, and local authorities campaigned to promote breast feeding, primarily, but also to provide subsidized milk to mothers with new children (Fildes, Marks, and Marland 1992). There were also increasing demands to improve further the hygiene of the dairy industry and milk supply. Bacteriologists had been studying the flora of milk to investigate its keeping properties and the transmission of infectious disease from animals to humans, with special attention to tuberculosis (TB) and the question of whether bovine TB could be passed to infants in milk. In Britain, the problem was thoroughly examined in a series of royal commissions on TB that effectively created a semipermanent body of bacteriologists investigating questions of milk and meat hygiene. In 1914, the Milk and Dairies Act was passed to prevent the sale of milk from tuberculous cows, while giving local authorities considerable powers of inspection (Bryder 1988; Smith 1988; Atkins 1992).

Similar measures were pursued in many other countries. In Australia, milk institutes were established in Brisbane and Melbourne in 1908 to ensure clean milk supplies (Mein Smith 1992). The same year saw the Canadian Medical Association devise a system of standards for milk hygiene in the dairy industry, which was implemented by local authorities. Toronto and Hamilton established pure milk depots before

World War I (Comacchio 1992). Free or subsidized milk was available in the Netherlands from *Gouttes de lait* - modeled closely on the pioneering French milk depots (Marland 1992). In American cities, the emphasis was more on ensuring the quality of milk supplies than on providing subsidized milk, and schemes for regulating standards were devised by many cities, including New York, Philadelphia, and Memphis, and the state of Illinois, with particular emphasis on eradicating TB and promoting the pasteurization of milk (Helper 1986; Shoemaker 1986; Meckel 1990; Peagram 1991).

Such attention given to infant welfare was incorporated into the wider schemes of social welfare developing in the early twentieth century (Lewis 1993). Enthusiasm for the provision of free milk seems to have been relatively short-lived and was replaced with an increasing emphasis on the education of mothers on matters of child care and good house-keeping. But the public-health principles underlying campaigns for clean milk persisted, and the ever-increasing volumes of liquid milk coming into the towns were carefully scrutinized.

Milk in the Twentieth Century

The twentieth century was marked by a massive expansion in the production and consumption of dairy commodities, particularly liquid milk. Rationalized, scientific dairy farming, based on calculated feeding, careful milk recording, and artificial insemination, has produced cows that are virtually milk machines. The production, manufacturing, and distribution of dairy products has become ever more concentrated, yet the continual growth of dairying has required equally prodigious efforts to find markets for the produce.

Around midcentury, milk was viewed as one of humankind's most important foods. Advertisers and scientists alike had convinced the public that milk was a supremely healthy, nourishing, even life-giving substance. Its almost mystical whiteness was, for perhaps the only time in its history, matched by its hygienic purity and popular appeal. Toward the end of the twentieth century, however, health problems connected with milk drinking were uncovered, and inappropriate Western marketing schemes involving milk were exposed as highly detrimental to infants' health in developing countries.

In the early twentieth century, however, dairying appeared to have a glorious future and, with ever-expanding urban populations, more and more farmers turned to liquid milk production. For particularly hard-pressed British farmers, dairying had become the cornerstone of agriculture, outstripping the economic return from cereals (Taylor 1974). In the United States, too, dairying was a fallback for farmers on relatively low-yielding land and a common means of diversifying a narrowly based state agriculture

(Cochrane 1979; Danbom 1989). Such a recourse to dairying, however, carried with it the danger of overproduction. Several strategies were pursued in an effort to deal with dairy surpluses, but by far the favored means was to promote demand. This had expanded on its own in prosperous times as increasing numbers of people with surplus income were able to extend the scope of their diets to include more fresh meat, vegetables, and dairy produce. Beginning in the 1920s and 1930s, a concerted effort was made to encourage consumption of dairy goods and, especially in Britain, the drinking of milk. Such an effort, however, required that the commodity be safe.

Experiments with various types of heat treatment had been conducted to extend the keeping properties of milk (Dwork 1987a). The principle had been established by Louis Pasteur in his studies of beer and wine, and bacteriologists and inventors were not slow to follow up on it. Well into the twentieth century, however, there was significant resistance to the heat treatment of milk (Pyke 1968). It was commonly believed that heating destroyed most of milk's essentially nutritious and health-giving properties, even a vitalist life force (McKee, personal communication). At the same time, there was a school of opinion within the dairy industry that a certain level of natural contamination was necessary to make milk into cheese or butter, and that pasteurized milk was unsuitable for dairy manufacture (Davis 1983). Although the arguments for pasteurization were accepted more readily in the United States than in Britain, debate continued into the 1930s (Meckel 1990). Opponents argued that heat treatment was a technical fix for sloppy procedures and an excuse to produce unclean milk (Davis 1983). Ultimately, however, an increasingly concentrated dairy industry cut through the controversy by routinely pasteurizing all milk it received, which was becoming a necessity because of the large volumes handled and the long distances now covered by milk distributors (Whetham 1976).

A considerable amount of research on nutrition and health was done during the interwar years, particularly in depressed areas of Britain. Experiments by Corry Mann and Boyd Orr on school children showed that groups given milk for a certain period put on more weight, had higher hemoglobin blood counts, and, according to their teachers, demonstrated a greater attentiveness at school; their parents commented that they were also livelier at home (Burnett 1989). In the controversial debates about diet in depressed Britain, milk was an important factor. It was thought that although welfare could not provide a sufficient quantity of food to maintain good health, this goal might be achieved with food supplements, such as milk, needed in fairly small quantities (Webster 1982).

By the 1930s, a culture of dairy consumption was well established in the major dairy manufacturing countries. On the Continent, this primarily involved

manufactured butter and, to a lesser extent, cheese; in the United States and in Scandinavia, where liquid milk consumption had a high profile, there was a sound base upon which marketing campaigns could expand demand (Teuteberg 1992). In Britain, although dairying was a major sector of agriculture (and one in which liquid milk was particularly prevalent), people neither ate as much butter as continental Europeans nor drank as much milk as Americans and Scandinavians. Milk was chiefly something to put in tea or give to children, old people, or invalids; indeed, throughout Europe, liquid milk was associated with sickness or effeminacy (McKee, personal communication).

Thus, one task of the Milk Marketing Boards for England, Scotland, and Northern Ireland, set up after the act of 1933, was to overcome prejudices against milk and stimulate demand for the products of an industry now suffering the consequences of overproduction from the decreasing prices that manufacturers were paying for milk. The Milk Marketing Boards (MMBs), as producer organizations, tried to regulate the dairy market like benign monopolists. Farmers were paid a fixed price for their milk, which was in turn sold to dairy product manufacturers at one price and to liquid milk distributors at a higher one, thus evening out the differences in income that depended on farmers' access (or nonaccess) to a liquid market. As the MMBs undertook to buy all the milk produced by dairy farmers, they had to find markets for it. From the start, a tariff was levied on each gallon to pay for advertising, and a massive campaign was launched, following American models. It stressed the value of milk for health, strength, and beauty, and featured sportsmen and women, manual laborers, and film stars (Jenkins 1970; Whetstone 1970; Baker 1973).

The emphasis on naturalness, whiteness, and purity reflected the contemporary concern for light, space, and healthy living. Milk was marketed to youth through futuristic, American-style milk bars – all chrome, steel, and modernity. The marketing of ice cream was another area in which the British copied from America and the Continent in trying to improve sales (McKee, personal communication). In 1934, drawing on recent discoveries of the vitamin content of milk, the company Cadbury launched a new advertising campaign for its dairy milk chocolate, which was marketed as a food as well as a treat (Othick 1976; Horrocks 1993).

Throughout the 1930s, the virtually unique retail system of doorstep deliveries was established in Britain. Derived from the roundsmen of the eighteenth century who pushed churn-carrying carts through the streets, milk distributors between the wars developed the service of daily deliveries of pint bottles of pasteurized milk to the doorstep (Baker 1973). The MMBs promoted the system and, until recently, the image of the milkman doing his rounds and the bottle on the doorstep were central to the

culture of milk in Britain, elevating the milkman to the status of folk hero or comic character. Only with the 1980s or so was there a notable decline in the importance of milkmen, as the greater prevalence of automobiles, refrigerators, and supermarkets have resulted in British consumers joining Americans and other Europeans in buying cartons of milk from stores (Jenkins 1970).

Milk remained important for children during the 1930s, with the Milk Act of 1934 providing subsidized milk for British schoolchildren. The act also continued the practice of baby clinics begun at the turn of the century, which encouraged the habit of drinking milk at an early age (Hurt 1985). Milk for pregnant women and new mothers was also a priority, and the welfare state distributed nourishment for mothers, as well as infants and children, until the program was discontinued in 1972 by the British parliament. The efforts of the MMBs were significant – so much so that as the 1930s came to a close, British people were drinking, on average, a pint of milk more per week than they had in the mid-1920s. Moreover, such a trend continued through the 1950s, when British per capita milk consumption exceeded that of the United States; only Sweden and Ireland had higher levels (Jenkins 1970).

In the last decades of the twentieth century, dairying and dairy products continued to hold an ambiguous status. The market for milk in the major dairy areas of the developed world seemed to be saturated and was even showing signs of diminishing. Consumption peaked sometime during the 1960s and then, depending on place, stabilized or began to shrink (OECD 1976). Nonetheless, dairying has continued to be a significant factor in industrial agriculture throughout the last half of the twentieth century. In Europe, the world's principal dairy-producing region, milk products accounted for between 11 and 35 percent of total farm sales (Butterwick and Rolfe 1968). Such regular cash income is a lifeline for farmers, but the vast sums paid out in subsidies have helped to swell the milk "lakes" and butter "mountains" of surplus production to the absurd point that butter has been used as a feed for the cows that produced the milk to begin with. Since the interwar period, dairy farming has been maintained only by subsidies and tariffs, with overproduction the unhappy outcome (Johnson 1973; Cannon 1987). Developing countries seem to be repeating the experience of the developed West, with liquid milk sectors emerging to supply urban areas (Kurien 1970; Chang 1977; Mahias 1988).

Although dairy farming has been sustained at considerable economic cost, other, more insidious problems have emerged for milk products. The high food value, which made it so important for infants and children until the mid-twentieth century, lead to more or less serious health problems in the overfed late twentieth century. The high-fat content of milk has been

implicated in the modern Western illnesses of obesity, coronary artery disease, and a host of digestive and allergic conditions. As a consequence, there has been a marked trend toward skimmed milk and milk powders, and even toward the exclusion of milk from the diet. Similarly, butter has been a victim of concerns about heart disease and body weight and has been steadily replaced by an increasingly more palatable margarine during the twentieth century (Cannon 1987).

To combat health worries, marketers of milk and milk products have focused on their energy-giving properties and high calcium content. Although milk could not realistically be promoted as a slimming aid, as it had been in the 1930s, the industry was still aiming for a pint per person per day, and advertised milk as a provider of instant vitality (Jenkins 1970). Some dairy marketers in recent years have undermined the ground of critics by deliberately proclaiming the richness and luxury of specialized products, such as fresh cream or ice cream. In an ironic reversal of healthy eating advice, cream has been lauded as "naughty – but nice!" and recent advertisements have been aimed at subverting any associations of ice cream with notions of childhood or purity.

A rather more sinister turn of events, however, has also tarnished the pure-white reputation of milk. In the early 1970s, the issue of inappropriate marketing of canned milk and milk powders among developing-world peoples began to receive public attention. In Yemen, for example, a study showed that dried milk and canned milk, provided as part of general food aid, were being fed to babies. This occurred in a country where Western medicine and high-technology goods had high social cachet and where the natural capacities of women were commonly denigrated. The result of the aid was that women neglected to breast-feed their babies and turned instead to proprietary baby foods or artificial milk. Many of these women were unable to read the instructions and had no access to the necessary sterilizing equipment to keep bottles clean, let alone to an adequate supply of fresh, clean water, with the result being an increase in infant mortality. Although resolutions have been made by dairy manufacturers in accordance with World Health Organization (WHO) and United Nations International Children's Emergency Fund (UNICEF) recommendations not to promote goods that might be used as substitutes for breast milk, contraventions continue (Melrose 1981).

Milk remains a singularly, almost uniquely, nutritious foodstuff, invested with elementally significant cultural and nutritional value. Yet milk has a Janus-faced nature – as it nourishes, so can it harm. Its whiteness evokes purity, but may also conceal corruption. Thus, as a commodity made widely available, milk has sometimes attacked the health it was intended to support and killed infants it was meant to nurture.

Keith Vernon

Bibliography

- Atkins, P. J. 1980. The retail milk trade in London c. 1790–1914. *Economic History Review* 33: 522–37.
1991. Sophistication detected: Or the adulteration of the milk supply, 1880–1914. *Social History* 16: 317–39.
1992. White poison? The social consequences of milk consumption, 1850–1930. *Social History of Medicine* 5: 207–27.
- Baker, S. 1973. *Milk to market. Forty years of milk marketing*. London.
- Bateman, F. 1990. The marketable surplus in northern dairy farming: New evidence by size of farm in 1860. *Agricultural History* 52: 345–63.
- Bourke, J. 1990. Dairywomen and affectionate wives: Women in the Irish dairy industry, 1890–1914. *Agricultural History Review* 38: 149–64.
- Bryder, L. 1988. *Below the magic mountain: A social history of tuberculosis in twentieth century Britain*. Oxford.
- Burnett, J. 1989. *Plenty and want. A social history of food in England from 1815 to the present day*. Third edition. London.
- Butterwick, M., and E. N. Rolfe. 1968. *Food, farming and the common market*. London.
- Cannon, G. 1987. *The politics of food*. London.
- Cassedy, J. H. 1991. *Medicine in America. A short history*. Baltimore, Md.
- Chang, K. C., ed. 1977. *Food in Chinese culture. Anthropological and historical perspectives*. New Haven, Conn.
- Cochrane, W. W. 1979. *The development of American agriculture. An historical analysis*. Minneapolis, Minn.
- Cohen, M. 1984. The decline of women in Canadian dairying. *Social History [Canada]* 17: 307–34.
- Comacchio, C. P. 1992. “The infant soldier”: Early child welfare efforts in Ontario. In *Women and children first. International maternal and infant welfare 1870–1945*, ed. V. Fildes, L. Marks, and H. Marland, 97–120. London.
- Cullen, L. M. 1992. Comparative aspects of Irish diet, 1550–1850. In *European food history*, ed. H. J. Teuteberg, 45–55. Leicester, England.
- Danbom, D. B. 1989. The North Dakota agricultural experiment station and the struggle to create a dairy state. *Agricultural History* 63: 174–86.
- Davenport-Hines, R. P. T., and J. Slinn. 1992. *Glaxo: A history to 1962*. Cambridge.
- Davis, J. G. 1983. Personal recollections of developments in dairy bacteriology over the last fifty years. *Journal of Applied Bacteriology* 55: 1–12.
- Dicksen, P. 1973. *The great American ice-cream book*. New York.
- Drummond, J. C., and A. Wilbraham. 1939. *The Englishman's food. A history of five centuries of English diet*. London.
- Dupré, R. 1990. Regulating the Quebec dairy industry, 1905–1921: Peeling off the Joseph label. *Journal of Economic History* 50: 339–48.
- Dwork, D. 1987a. *War is good for babies and other young children. A history of the infant and child welfare movement in England, 1898–1918*. London.
- 1987b. The milk option. An aspect of the history of the infant welfare movement in England 1898–1908. *Medical History* 31: 51–69.
- Fildes, V. A. 1986. *Breasts, bottles and babies. A history of infant feeding*. Edinburgh.
- Fildes, V., L. Marks, and H. Marland, eds. 1992. *Women and children first. International maternal and infant welfare 1870–1945*. London.
- Finlay, M. R. 1990. The industrial utilization of farm products and by-products: The U.S.D.A. regional research laboratories. *Agricultural History* 64: 41–52.
- Fussell, G. E. 1926–9. The London cheesemongers of the eighteenth century. *Economic Journal* (Economic History Supplement) 1: 394–8.
1966. *The English dairy farmer 1500–1900*. London.
- Fussell, G. E., and C. Goodman. 1934–7. The eighteenth century traffic in milk products. *Economic Journal* (Historical Supplement) 3: 380–7.
- Harris, M. 1986. *Good to eat. Riddles of food and culture*. London.
- Haystead, L., and G. C. File. 1955. *The agricultural regions of the United States*. London.
- Helper, R. W. 1986. Bovine T.B. and the battle for pure milk in Memphis, 1910–11. *West Tennessee Historical Society Papers* 40: 6–23.
- Horrocks, S. 1993. Consuming Science: Science, technology and food in Britain, 1870–1939. Ph.D. thesis, University of Manchester.
- Hurt, J. 1985. Feeding the hungry schoolchild in the first half of the twentieth century. In *Diet and health in modern Britain*, ed. D. Oddy and D. Miller, 178–206. London.
- Jenkins, A. 1970. *Drink a pint. The story of milk and the industry that serves it*. London.
- Jensen, J. M. 1988. Butter making and economic development in mid-Atlantic America from 1750–1850. *Signs* 13: 813–29.
- Johnson, D. G. 1973. *World agriculture in disarray*. London.
- Johnson, K. 1971. Iowa dairying at the turn of the century: The new agriculture and progressivism. *Agricultural History* 45: 95–110.
- Keillor, S. J. 1993. Agricultural change and crosscultural exchange: Danes, Americans and dairying, 1880–1930. *Agricultural History* 67: 58–79.
- Kurien, V. 1970. The Anand and Bombay milk projects. In *Change in agriculture*, ed. A. H. Bunting, 199–206. London.
- Lampard, E. E. 1963. *The rise of the dairy industry in Wisconsin*. Madison, Wis.
- Lewis, J., ed. 1993. *Women and social policies in Europe*. London.
- Mahias, M.-C. 1988. Milk and its transmutations in Indian society. *Food and Foodways* 2: 265–88.
- Marland, H. 1992. The medicalisation of motherhood: Doctors and infant welfare in the Netherlands, 1901–1930. In *Women and children first. International maternal and infant welfare 1870–1945*, ed. V. Fildes, L. Marks, and H. Marland, 74–96. London.
- McKeown, T. 1969. *The modern rise of population*. London.
- Meckel, R. A. 1990. *Save the babies. American public health reform and the prevention of infant mortality 1850–1929*. Baltimore, Md.
- Mein Smith, P. 1992. “That welfare warfare”: Sectarianism in infant welfare in Australia, 1918–1939. In *Women and children first. International maternal and infant welfare 1870–1945*, ed. V. Fildes, L. Marks, and H. Marland, 230–56. London.
- Melrose, D. 1981. *The great health robbery. Baby milk and medicines in Yemen*. Oxford.
- Melvin, P. A. 1983. Milk to motherhood: The New York milk committee and the beginning of well-child programs. *Mid-America* 65: 111–34.
- Mennell, S. 1985. *All manners of food. Eating and taste in England and France from the Middle Ages to the present*. Oxford.

- Murray, J. W. 1977. *Growth and change in Danish agriculture*. London.
- Nurnally, P. 1989. From churns to "butter factories." The industrialization of Iowa's dairying, 1860-1900. *Annals of Iowa* 49: 555-69.
- Oddy, D., and D. Miller. 1976. *The making of the modern British diet*. London.
1985. *Diet and health in modern Britain*. London.
- OECD (Organization for Economic Co-operation and Development). 1976. *Study of trends in world supply and demand of major agricultural commodities*. Paris.
- Offer, A. 1989. *The First World War: An agrarian interpretation*. Oxford.
- O'Grada, C. 1977. The beginnings of the Irish creamery system 1880-1914. *Economic History Review* 30: 284-305.
- Okun, M. 1986. *Fair play in the marketplace. The first battle for pure food and drugs*. DeKalb, Ill.
- Osternd, N. G. 1988. The valuation of women's work: Gender and the market in a dairy farming community during the late nineteenth century. *Frontiers* 10: 18-24.
- Othick, J. 1976. The cocoa and chocolate industry in the nineteenth century. In *The making of the modern British diet*, ed. D. Oddy and D. Miller, 77-90. London.
- Peagram, T. R. 1991. Public health and progressive dairying in Illinois. *Agricultural History* 65: 36-50.
- Petty, C. 1989. Primary research and public health: The prioritization of nutrition research in interwar Britain. In *Historical perspectives on the MRC*, ed. J. Austoker and L. Bryder, 83-108. London.
- Pyke, M. 1968. *Food and society*. London.
- Rosenberg, C. E. 1976. *No other gods. On science and American social thought*. Baltimore, Md.
- Shoemaker, S. T. 1986. The Philadelphia pediatric society and its milk commission 1896-1912: An aspect of urban progressive reform. *Pennsylvania History* 53: 273-88.
- Smith, F. B. 1988. *The retreat of tuberculosis, 1850-1950*. London.
- Smith, R. E. F., and D. Christian. 1984. *Bread and salt. A social and economic history of food and drink in Russia*. Cambridge.
- Tannahill, R. 1973. *Food in history*. New York.
- Taylor, D. 1971. London's milk supply, 1850-1900: A re-interpretation. *Agricultural History* 45: 33-8.
1974. The English dairy industry, 1860-1930: The need for a reassessment. *Agricultural History Review* 22: 153-9.
1976. The English dairy industry, 1860-1930. *Economic History Review* 29: 585-601.
1987. Growth and structural change in the English dairy industry, c. 1860-1930. *Agricultural History Review* 35: 47-64.
- Teuteberg, H. J. 1982. Food consumption in Germany since the beginning of industrialization. A quantitative, longitudinal approach. In *Consumer behaviour and economic growth in the modern economy*, ed. H. Baudet and H. van der Meulin, 233-77. London.
- Teuteberg, H. J., ed. 1992. *European food history. A research review*. Leicester, England.
- Tweeddale, G. 1990. *At the sign of the plough. 275 years of Allen and Hanbury's and the British pharmaceutical industry 1715-1990*. London.
- Valenze, D. 1991. The art of women and the business of men: Women's work and the dairy industry c. 1740-1840. *Past and Present* 130: 142-69.
- Warner, M. 1976. *Alone of all her sex. The myth and cult of the virgin Mary*. London.

Webster, C. 1982. Healthy or hungry thirties? *History Workshop Journal* 13: 110-29.

Whetham, E. H. 1964. The London milk trade, 1860-1900. *Economic History Review* 17: 369-80.

1976. The London milk trade, 1900-1930. In *The making of the modern British diet*, ed. D. Oddy and D. Miller, 65-76. London.

Whetstone, L. 1970. *The marketing of milk*. London.

Yudkin, J., ed. 1978. *Diet of man: Needs and wants*. London.

III.10 ☞ Soft Drinks

In considering the human body's demand for food and nutrition, the simple need for liquid refreshment is sometimes overlooked. Although this fundamental physiological requirement can be satisfied by drinking an adequate supply of pure water, most people, when given a choice, prefer to achieve the required level of liquid intake with a variety of flavored drinks to stimulate the palate.

Soft drinks are usually defined as nonalcoholic, water-based drinks, although a few may contain alcohol, albeit in quantities too small to warrant their classification as "hard liquor." Soft drinks are usually sweetened - soda water being an obvious exception - and flavored with food acids, essences, and sometimes fruit juices. They are often carbonated - that is, charged with carbon dioxide gas - and, indeed, in North America are referred to as carbonated beverages. In some countries, including the United Kingdom, there is a significant retail market for concentrated soft drinks intended for dilution at home before consumption. Soft drinks in powdered form are similarly marketed for preparation at home. In addition, uncarbonated, ready-to-drink soft drinks are also found.

The flavors of soft drinks may be derived from fruits, nuts, berries, roots, herbs, and other plants. Moreover, fruit (and to some extent vegetable) juices, as such, have grown in popularity in recent years and have come to be included among the soft drinks. In many countries, soft drinks are distinguished from hard liquor by the higher taxation of stronger drinks, for example through excise duties, and the term "non-alcoholic" can sometimes mean merely "non-excisable." However, soft drinks are frequently subject to other taxes, though usually at lower levels than those that are levied on alcoholic drinks. Soft drinks are often distinguished from medicines by legislation. In the past, these distinctions were less precise, and a historical study of soft drinks will include products that began with a mainly medicinal purpose but are regarded today as simple refreshment.

Over the years, the names of various classes of soft drinks have been used very imprecisely. The term "mineral waters," originally and properly confined to spa and spring waters, has subsequently been used

for artificial spa waters and even for flavored carbonated drinks. Even today, words like “lemonade” or “pop” may be used colloquially to embrace a wide range of drinks – an imprecision that often prevents our determining the flavor or, indeed, the composition of past drinks. It cannot be assumed, merely because a bygone drink bore the same name as a current one, that it necessarily shared the same compositional standards. Another generic term, “aerated waters,” is believed to have been coined by the eighteenth-century French chemist Gabriel Venel, who, in preparing an artificially carbonated water, called it “eau aérée” (aerated water), erroneously believing the carbon dioxide gas to be mere air.

The commercial manufacture of prepacked soft drinks began in the last years of the eighteenth century, and some of its products were known in one form or another well before that. The term “soft drink,” however, dates only from the last years of the nineteenth century and seems to have originated in the United States. It is, therefore, strictly anachronistic to refer to such bygone products as soft drinks, although admittedly, an appropriate alternate term is lacking.

Small Beer

Before the mid-seventeenth century, the principal European drinks were what today would be considered alcoholic: beer, ale, mead, cider, perry (fermented pear juice), and wine. Until that time, and beyond, there was – certainly in England – a distinct prejudice against drinking water, as such, unless from sources of proven reputation. Nor was this prejudice wholly unjustified given the contamination of much of the water supply, particularly in populous districts. Nonetheless, cheaper and weaker ales (“small ales”) and beers (“small beers”), containing insignificant amounts of alcohol, were produced for those who could afford nothing stronger, and for children. In his fourteenth-century *Piers the Plowman*, William Langland refers to halfpenny ale and penny-ale as beyond the pockets of the poorest laborers. In Leicester, the brewers were especially enjoined to make “good wholesome small drink for the poor people” (Bateson 1901, 2: 288).

An English cookery book of the fifteenth century contains a list of “herbs for the cup.” Among those to be grown in the garden were sage, rosemary, hyssop, marjoram, and gillyflower; these were as likely to be used for flavoring mead (made from honey) as for small beers. Nettle beer and heather ale were also found among the small beers. Not surprisingly, small beer was despised by the more hardened drinker. “Doth it not show vilely in me to desire small beer?” asked Shakespeare’s Prince Hal before his reformation (*Henry IV, Part II* 2.2.7). And, in mentioning Shakespeare, it is humbling to note that for him, “to chronicle small beer” was to deal in very minor matters indeed!

The strength of church ales is less certain. They were drunk as both a social custom and as a fund-raising exercise. In the seventeenth century, John Aubrey recalled the church ales of his youth as a means of raising funds for the poor before the introduction of a formal system of taxation by local government: “In every parish is, or was, a church house to which belonged spits, crocks etc., utensils for dressing provision. Here the housekeepers met, and were merry and gave their charity: the young people came there too, and had dancing, bowling, shooting at butts, etc., the ancients sitting gravely by, looking on. All things were civil and without scandal” (Barber 1988: 184).

Church ales were numerous but their names distinguished them more by use than by strength. Among them were bride ale, wake ale, and Whitsun ale, all intended for fund-raising of various sorts. Aubrey noted that “the clerk’s ale was in the Easter holidays, for his benefit” (Barber 1988: 184). These ales were probably stronger than the small beers; despite Aubrey’s assertion of their innocent intent, they certainly attracted clerical and Puritan criticism as giving rise to licentiousness and disorder.

Small beer, however, survived the years. It was to be found in the coffeehouses of late Stuart and Georgian London. In the early nineteenth century, William Cobbett noted that the grass mowers’ drink allowance was “two quarts of what they call strong beer, and as much small beer as they can drink” (Jekyll and Jones 1939: 112). An inventory of the cellars of a Berkshire squire at his death in 1822 showed 210 gallons of small beer out of 2,630 gallons of beer and ale of all sorts. Small beer, indeed, but still a measurable quantity.

In his 1833 report *The Poor Laws in London and Berkshire*, Edwin Chadwick noted that every convict on board hulks in England was allowed one pint of small beer a day as part of his diet, and Dr. Jonathan Pereira’s *Treatise on Food and Diet* of 1843 included small beer – or, as he called it, table beer – among his dietaries of London hospitals and other institutions. Pereira incidentally noted that whereas a barrel of best Burton ale contained 40 to 43 pounds weight more than an equivalent barrel of water, a barrel of good table beer contained 12 to 14 pounds more, and “common table beer” but 6 pounds more than water. Although in the 1790s Dr. Erasmus Darwin’s *Zoonomia* had recommended small beer as part of a diet against “gaol [jail] fever” (typhus), we may suspect that the drink was preferred by these institutions as much for economy as for health.

Other brews were not necessarily so weak. Spruce beer, flavored with leaves of the spruce fir, was known to sailors in the Baltic from at least the sixteenth century, and Canadian Indians taught Jacques Cartier to use it against scurvy at much the same time. In New Zealand in 1773, the explorer Captain James Cook brewed spruce beer and reckoned it a useful defense against the same disease. Unlike most small ales and small beers, which continued to be

brewed as and when required rather than being bottled for trading, spruce beer made the transition to the commercial market for prepacked soft drinks, and "spruce beer manufacturer" was to be found as an entry in the London trade directories of the early nineteenth century. And whatever its earlier strength may have been, spruce beer was excluded by the Licensing Act of 1872 from its definition of intoxicating liquor.

Ginger beer and ginger ale are products of the commercial era, but their acceptability, no doubt, owed much to the tradition of small beers and herbal brewing, which continued in the production of hop bitters and the like.

Cordials and Other Domestic Drinks

As small beers derived from the arts of brewing, so cordials owed their origins to the secrets of distillation. Heavily sweetened and highly flavored so that they might even be diluted with water before drinking, cordials would vary in alcohol content according to the recipe of their maker, often a well-to-do country housewife, whose object – partly pleasurable, partly medicinal, but at all times designed to tempt the palate – was summed up in Shakespeare's phrase, "a taste as sweet as any cordial comfort" (*Winter's Tale* 5.3.77).

Homemade cordials survived well into the era of commercial soft drinks, and as late as 1856, George Dodd in *The Food of London* described them as "more frequently the handiwork of some Lady Bountiful, some housewife more than ordinarily clever in domestic economy, than of manufacturers who prepare them for sale" (Dodd 1856: 498). Nonetheless, by then, cordials were also available from commercial manufacturers in both alcoholic and nonalcoholic varieties, the latter being popular as temperance drinks, and they continued in essence-based peppermint, ginger, and clove cordials, the thought of their medicinal origins having for the most part faded.

Among other drinks from the domestic sickroom was barley water, an infusion of pearl barley and water dating from late medieval times, which Thomas Fuller in the seventeenth century described as "an invention which found out itself, with little more than the bare joining the ingredients together" (Fuller 1662: 366). For sixteenth-century invalids there was "water imperial," apparently containing sugar and cream of tartar and flavored with lemons, as well as "manays cryste," a sweetened cordial flavored with rosewater, violets, or cinnamon.

Fruit-Flavored Drinks

In view of the heavy Arab influence on Italian Renaissance cuisine, lemonade may have originated with the Arabs. But in any event, sixteenth-century Italians seem to have been the first Europeans to enjoy this

beverage made from freshly squeezed lemons, sweetened with sugar or honey, and diluted with water to make a still, soft drink that could be prepared, sold, and consumed on the premises. Its popularity spread to France and gradually to the rest of Europe, until by the eighteenth century, lemonade of this sort was available from the inns of Scotland to the Turkish baths of Constantinople. In France, lemonade was sold by the itinerant *limonadier*, who stored the drink in a tank carried on his back. In 1676 the *limonadiers* of Paris were formed into a company and granted a patent or monopoly by the government, continuing to sell their drink in this way until at least the end of the following century. During the eighteenth century, lemonade was also valued by the medical profession, and Erasmus Darwin recommended it, among other things, for the relief of kidney stones and gout and in cases of scarlet fever.

Orange juice was first introduced into mid-seventeenth-century England; Samuel Pepys in the 1660s noted with approval this drink that was new to him. A little later, orangeade was also to be found, often containing oranges too bitter to be eaten fresh. Orgeat, a cooling drink flavored with almonds and orange flower water, became a favorite of the patrons of eighteenth-century London refreshment houses and pleasure gardens.

Lemons and Scurvy

In the eighteenth century and earlier, citrus juices were among many articles of diet used in attempts to find a cure for scurvy, a disease which only in the twentieth century was discovered to result from a dietary deficiency of vitamin C. Scurvy particularly affected sailors on extended voyages of discovery with few opportunities for revictualing with fresh foods. Until its cause was known, any cure could only be found by empirical tests. Beer brewed from the spruce fir was considered efficacious. Unsuccessful, however, was malt, although it too had its advocates for a time.

Lemon juice was favored by the early Spanish explorers as an antiscorbutic, and Dutch and English voyagers also included it in their ships' stores, although it was more likely to find a place among the medicines than as a regular article of diet. In the mid-eighteenth century, James Lind conducted and published the results of experiments at sea. His 1753 *A treatise on the scurvy* showed that sailors treated with lemon juice recovered from scurvy, whereas other sailors given other substances did not. But further practical tests were less conclusive, almost certainly because of the loss of vitamin C during the preparation and storage of the juice.

For instance, Captain Cook on his voyages to the Pacific was supplied with lemon juice as a concentrated syrup, with most of the vitamin C unwittingly boiled out in the preparation. Not surprisingly, he was

unenthusiastic about the efficacy of citrus juices, even though Joseph Banks, the botanist on the voyage with Cook, successfully dosed himself with lemon juice against what appeared to be the onset of scurvy. It was not until the end of the eighteenth century that the British Admiralty Board introduced lemon juice into the seaman's diet, where it was usually preserved by mixing with rum. In the mid-nineteenth century, lemon juice was largely replaced on British ships with West Indian lime juice. Botanical differences between lemons and limes were little appreciated at that time, and in fact, lime juice, with its lower levels of vitamin C, was less suited to the purpose (Carpenter 1986).

Artificial Mineral Waters

In 1772, Dr. Joseph Priestley's *Directions for Impregnating Water with Fixed Air* (fixed air being his name for carbon dioxide gas) also excited the interest of those seeking a reliable antiscorbutic. Priestley was by no means the first scientist to interest himself in the possibility of artificially reproducing the properties of natural mineral waters. As Sir John Pringle put it when Priestley received the gold medal of the Royal Society in 1773:

Having learned from Dr. Black that this fixed or mephitic air could in great abundance be procured from chalk by means of diluted spirits of vitriol; from Dr. Macbride that this fluid was of a considerable antiseptic nature; from Mr. Cavendish that it could in a large quantity be absorbed in water; and from Dr. Brownrigg that it was this very air which gave the briskness and chief virtues of the Spa and Pyrmont waters; Dr. Priestley . . . conceived that common water impregnated with this fluid alone, might be useful in medicine, particularly for sailors on long voyages, for curing or preventing the sea scurvy (Pringle 1774: 15).

Intellectual curiosity, rather than commercial advantage, seems to have prompted the early scientists to find ways of extracting and analyzing the salts from natural mineral waters and reconstituting them in their laboratories. Much the same spirit led Priestley to show how water might be artificially carbonated on a commercial scale, and although carbonated water proved no cure for scurvy, Priestley's invention was soon adapted to the commercial production of artificial mineral waters.

Despite the traditional widespread suspicion of water because of its close connection with disease, natural waters were nonetheless valued so long as they either contained mineral salts found in practice to be healthful or were drawn from an exceptionally pure and reliable source. Such waters, however, had to be highly regarded indeed, in light of the high cost of transporting them, usually in heavy glass bottles, over any but short distances. During the reign of George II,

Henry Eyre of London was not only importing from the Low Countries the mineral waters of Spa and dealing in various native waters, but he was also ensuring that all bottles were appropriately sealed to protect his customers from spurious imitations.

Provided the artificial waters carefully replicated the chemical composition of their natural counterparts – which analytical techniques enabled them to do – the economic advantage of manufacture close to the consumer was obvious. As Priestley himself put it: "I can make better than you import; and what cost you five shillings, will not cost me a penny" (Rutt 1831, 1: 177). At much the same time, Torbern Bergman of Uppsala, Sweden, was also experimenting with equipment for the production of artificial mineral waters, and within the next decade, Dr. John Mervyn Nooth was demonstrating to the Royal Society in London a glass apparatus for the production of small quantities of carbonated water.

The first known manufacturer of artificial mineral waters bottled for sale was Thomas Henry, a Manchester apothecary who, by the end of the 1770s, had modified Nooth's apparatus in order to produce artificial Pyrmont and Seltzer waters, as well as to imitate an earlier preparation known as "Bewley's Mephitic Julep," all of which were intended for medicinal purposes rather than for refreshment. Indeed, Henry recommended drinking with the julep "a draught of lemonade, or water acidulated with vinegar or weak spirits of vitriol, by which means the fixed air will be extricated in the stomach" (Henry 1781: 29). This suggestion reflected earlier advice on taking the natural waters; there was no hint, as yet, that flavorings might be added to the waters themselves.

J. H. de Magellan, claiming that Nooth's apparatus took several hours to impregnate water, published (1777) his own method for producing, in a few minutes, artificial versions of "the best Mineral Waters of Pyrmont, Spa, Seltzer, Seydschutz, Aix-la-Chapelle etc." (Magellan 1777: Title), as well as appropriate recipes for doing so. At the same time, he mentioned that he had sent copies of Priestley's pamphlet to different parts of Europe and that a French translation had appeared soon afterward. In 1787, mineral waters were also said to have been manufactured on a commercial scale in Germany.

It was, however, Jacob Schweppe who took up a theoretical suggestion of Priestley's that the use of a "condensing engine" or pressure pump would allow a greater volume of gas to be absorbed in the water than was otherwise possible. Schweppe, German born and a citizen of Geneva by adoption, pioneered the manufacture of artificial mineral waters in that city before setting up business in London in 1792. Producing Seltzer water, Spa water, Pyrmont water, and acidulous Rochelle salt water on something approaching a factory scale, Schweppe also offered a less specific line of aerated alkaline water, soon known as acidulous soda water, which he sold in three strengths – sin-

gle, double, and triple – according to the amount of soda present, the double being “generally used.”

The success of such ventures depended not merely on carbonating the water but on retaining the gas in the liquid until the consumer opened the container, and this in turn depended on the careful corking and sealing of all bottles, which Henry had stressed at the very outset of commercial manufacture. For Scheweppe, who was supplying sometimes over long distances, the problem was a real one. As a postscript to a repeat order of 1805: A Birmingham customer complained that many of the bottles from his last order were nearly empty when they arrived because of bad corking.

To prevent the corks from drying out, Scheweppe recommended that the bottles be laid on their sides in a cool place or even better kept covered with water – no easy task on a carrier’s wagon! Scheweppe also made an allowance on empty bottles returned, a custom often retained thereafter in a trade where the bottle represented a significant proportion of the total cost of the product.

Scheweppe’s partner in Geneva, Nicolas Paul, also made his way to England and operated commercially in London from 1802, having been in business in Paris for a while en route. Like Scheweppe, Paul used the process known as the Geneva system or Geneva apparatus, but although he apparently achieved even higher levels of carbonation than Scheweppe, the additional gas was, no doubt, largely lost in the pouring out.

By then, soda water had reached Dublin, where it was recommended by Dr. Robert Percival, Professor of Chemistry at Trinity College. Indeed, at one time it was claimed that a Dublin firm had invented soda water, but the product would seem conclusively to have originated with Scheweppe. Nonetheless, its early success in Britain, like that of the other artificial waters, was undoubtedly medicinal; in fact, between 1804 and 1833, soda water was subject to stamp duty under the Medicine Tax. The first known manufacturer of soda water in the United States was Benjamin Silliman, operating in New Haven in 1807, and the first United States patent for manufacturing artificial mineral water was issued two years later.

Attempts to match spa waters artificially probably reached their apogee in the 1820s, when Dr. F. A. Struve, of Dresden, opened a range of artificial spas at Leipzig, Hamburg, Berlin, St. Petersburg, and Brighton, supplying careful imitations of the Carlsbad, Ems, Kissingen, Marienbad, Pyrmont, Seltzer, and Spa waters to invalids and a wider public without the necessity of their traveling to the original waters’ respective sources.

Elegant and Refreshing Beverages

Gradually, however, the new drinks began to be promoted for refreshment rather than for their specifically medicinal properties. In 1819, an advertisement

in the London *Morning Chronicle* described as “elegant and refreshing” the ginger beer and soda water available from one of the new metropolitan makers, and it was by the brewing of ginger beer that the industry expanded from the soda waters, on the one hand, to the sweetened, fruit-flavored drinks of later manufacturers on the other hand.

In Elizabethan times, Arthur Barlowe’s “The discovery of Virginia” had referred to the American Indians drinking water “sodden [i.e. boiled] with ginger in it, and black cinnamon, and some times sassafras, and divers other wholesome and medicinable herbs and trees” (Barlowe 1589). A subsequent early American drink was “switchell,” a mixture of molasses, vinegar, and ginger. But no reference has been found to ginger beer as such before the first decade of the nineteenth century. After it was first marketed in England, however, its popularity grew swiftly. Perhaps, as Leigh Hunt wrote at the time, because it was found to have “all the pleasantness and usefulness of soda-water without striking cold upon one” (Hunt 1862), ginger beer soon became a staple commodity of even the most modest refreshment stall.

The commercial origins of flavored, sweetened carbonated waters remain more obscure. As early as 1784, Karl Wilhelm Scheele, the Swedish chemist, had produced, from lemon juice, a crystalline substance that he called citric acid. One old trade historian claimed to have seen a manuscript reference to citric acid powder (or concrete acid of lemons, as it was also known) dated 1819, and a recipe for lemonade made with citric acid dated shortly thereafter. Lemonade “syrup” was known at about the same time, and all of these substances may well have been used for making lemonade in the home, employing a variant of Nooth’s apparatus later known to the Victorians as the “gazogene” or “seltzogene.” Despite such speculation, however, the first positive reference in England to commercially produced effervescent lemonade dates from no earlier than 1833. Such lemonades would have been flavored with citric acid and essential oil of lemon mixed with a sugar syrup, topped up with water, and impregnated with carbon dioxide gas derived in the factory from the action of sulfuric acid on whiting or other forms of chalk. Only at the end of the century did carbon dioxide gas come to be supplied to the soft-drink factory by a specialist manufacturer.

At London’s Great Exhibition of 1851, lemonade, ginger beer, spruce beer, seltzer water (by now a generic name), and soda water were among the refreshments available to visitors, alcoholic drinks not being countenanced on the premises. A million bottles of aerated beverages were sold there, according to one contemporary estimate, and the success of the show and of the soft drinks reflected the slowly increasing leisure and spending power of those able to attend, two vital factors in the growth of the industry, not only in Britain but elsewhere in the developed world.

At fairs and race meetings, London costermongers sold lemonade and ginger beer, which they made at home in stone barrels. At the street markets of the mid-nineteenth century, according to Henry Mayhew (1851), soft drinks were available to refresh those shoppers "who have a penny to spare rather than those who have a penny to dine upon." Besides lemonade, raspberryade, and ginger beer, the street markets offered "Persian sherbet, and some highly coloured beverages which have no specific name, but are introduced to the public as 'cooling' drinks; hot elder cordial or wine; peppermint water." Sherbets had been available since at least the seventeenth century as cool fruit drinks originating from Turkey and the East, but later the name became attached to drinks made from effervescent powders containing bicarbonate of soda, tartaric acid or cream of tartar, sugar, and flavorings. As to the drinks with "no specific name," perhaps it is as well that their composition remains a mystery. In the street markets, Mayhew also noted that "some sellers dispensed ginger beer in plain glass bottles which was drank straight from the bottle after the cork obviated the necessity of a glass."

In France the development of the industry seems to have been slower, with pharmacists holding a monopoly on what remained a localized trade, until their grip was challenged and weakened during the Orleanist years of the 1830s and 1940s. From France, too, at that time, came the soda siphon for dispensing carbonated soft drinks.

The pharmacists of the United States also became adept at producing artificial mineral waters, and as they discovered that the installation of soda fountains brought customers to their retail drugstores, they were encouraged to experiment with an ever wider range of flavored drinks by the mid-nineteenth century. Soda fountains were also taken up in Europe, where, for example, Germany had its *Trinkballen* and France its *buvettes à eaux gazeuses*. An innovation of the American soda fountains was the addition of sweet cream to many of their products, and the popularity of ice-cream soda, as a vanilla-flavored drink, spread overseas as the century progressed.

Other drinks originating as soda fountain beverages included sarsaparilla, originally a medicinal flavoring derived from plants of the *smilax* species. Curiously, by the time sarsaparilla had become established among the bottled drinks, its flavor had come to be derived from a blend of oil of wintergreen, sassafras, anise, orange, and sometimes licorice. According to Charles Sulz in 1888, sarsaparilla itself was seldom included among the ingredients of what was by then a staple beverage of the American industry, but when it was, the bottler advertised its presence as proof of the superiority of his product over those of his competitors. Root beer was another blend of root, herb, and fruit flavors originating at the American soda fountain but later widely bottled. Sarsaparilla root beer, too, was available in the America of the

1880s. It also contained sassafras, a flavoring derived from the sassafras tree of the eastern seaboard.

Mixer Drinks

Freed from its exclusively medicinal limitations, soda water was found useful in diluting wines and spirits. Lord Byron called for hock and soda water after a drinking bout, and gin and soda water was said to be a favorite tipple of the fast set in the English hunting shires of the 1830s. Brandy and soda later became the drink for gentlemen, and manufactured seltzer and lithia waters were also available as mixers. As soda water became increasingly used as a mixer, many makers gradually reduced its soda content until, eventually, they were applying the name to simple carbonated water. The more scrupulous manufacturer sold such a product as "table water," and if his soda water retained a significant soda content, he was quick to advertise the fact.

Natural spa and spring waters, too, were used as mixers. Even in England, the natural waters of continental Europe were still imported despite local imitations. One London supplier in the 1860s offered not only German, French, and Belgian waters but others from Austro-Hungary and the United States.

The first English soft drink developed specifically as a mixer appears to have been tonic water, which began as a palatable means of ingesting the quinine prescribed for sufferers of malaria contracted in the tropics. In 1858 its inventor, Erasmus Bond, patented it as an "Improved Aerated Liquid," soon known as Indian or quinine tonic water.

Ginger ale seems to have originated a little later, once a method was devised for producing the clear extract of ginger, which distinguished the product from the cloudy ginger beers. A British trade paper of 1876 described the drink as a thing unknown only a few years previously. By then, however, it had become a favorite of the British market and was exported in significant quantities, particularly to America, and principally by the soft-drink firms of Belfast. Indeed, by the turn of the century a trade advertisement was asking: "What go-ahead Mineral Water Maker is there who has not at one time or another longed for the day to come when he would be able to turn out a Ginger Ale equal to the world-famous Belfast Ginger Ales?"

Temperance Drinks

Ginger ale, unadulterated, also took its place among the temperance beverages of Britain and the United States as the powerful social and religious forces moving against the consumption of alcohol gathered momentum on both sides of the Atlantic in the second half of the nineteenth century. Soft-drink manufacturers, serving licensed and temperance outlets alike, generally maintained a diplomatic neutrality in

the fierce battles over the drink trade, at the same time expressing occasional ironic amusement at the reformers' description of their nonalcoholic favorites by reference to the names of the very drinks they sought to defeat.

Hop ales, hop beers, dandelion stout: All these and more endeavored to provide alternatives to the workingman's beer, whereas football stout and football punch aimed to attract young men from the sports field. One winter punch was advertised as "the best non-alcoholic substitute for brandy," but in what respect was unspecified. Many such nonalcoholic beverages were fermented, but as one turn-of-the-century writer cryptically noted, "some of them are not fermented and others are not non-alcoholic." Uproar occasionally ensued when analysis revealed that a temperance drink contained as much, if not more, alcohol than the product it sought to supersede.

Further up the market, a full range of nonalcoholic champagnes – sweetened, flavored, carbonated drinks, usually of high quality – resembled champagne in their presentation but not their origins, while nonalcoholic fruit wines, drunk as such or diluted, imitated the syrupy consistency of liqueurs. The American soda fountain, too, flourished under the temperance movement, offering an ever-greater selection of flavors and blends of flavors. This vast increase in choice was the result not only of demand but of the growth of specialist essence houses that supplied the soft-drink industry with flavorings and with careful advice on how best to use them. As a result, essences even came to be used in the manufacture of ginger beer, but the result, although of more uniform consistency, was generally held to lack a certain something of the brewed original.

By then, however, much of the ginger beer available in England was of poor quality, hence this lament from the 1880s: "‘Times were’ when ginger beer *was* ginger beer, as the name implies; but *now* ’tis generally something quite different" (Good 1880). The writer, Joseph Gould (1880), went on to deplore the widely varying standards by which the drink was being made, in most of which ginger was "conspicuous by its absence," the product having become too often simply another sort of lemonade.

It was also toward the end of the nineteenth century that saccharin became available as an alternative sweetener to sugar in soft drinks. Discovered in 1879, the intensely sweet coal-tar derivative was patented for commercial manufacture in 1885, and early enthusiasts predicted a great future for it in soft-drink manufacture. In practice, although it became a particularly useful sweetener of drinks for diabetics (and later for low-calorie or diet drinks), when it was used in standard products critics considered its no match for the "body" or the palatability given by sugar, which over the years had come to be specially refined for soft-drink manufacture.

Fruit Drinks

The use of lime juice as an antiscorbutic for British seamen has already been noted and the Merchant Shipping Act of 1867 made its provision on shipboard a legal requirement. Until then the method of preserving lime juice had been in a mixture with 15 percent alcohol, but in that year, Lauchlan Rose, a lime and lemon merchant of Leith in Scotland, patented a means of preserving the juice without alcohol. Noting the method of preserving light wines by burning sulfur candles in the casks, Rose prepared a sulfur dioxide solution by passing the gas from burning sulfur through water. When this solution was added to fruit juice, it prevented fermentation and other defects to which unpreserved juices were liable when stored, and Rose marketed the result as lime juice cordial. On shore, the temperance movement and the mixer trade soon found additional uses for the new product.

In the late nineteenth century, squash came onto the market, originally as a still, cloudy, ready-to-drink, juice-based product. In the early twentieth century, John Dixon, an Australian manufacturer, began to put up concentrated fruit drinks, and concentrated lemon squash was introduced into Britain just before World War I. After the war, orange squash followed, then pineapple and grapefruit. Of these, orange and lemon squash proved the most popular.

The origins of barley water as a product of the kitchen have been noted in the section "Cordials and Other Domestic Drinks." The Victorians valued barley water as a drink for the sickroom and also as a temperance beverage, but it was still made at home from commercially available patent barley and, by then, often sweetened and flavored with lemon peel. It was not until the 1930s that bottled lemon barley water was successfully marketed as a concentrated soft drink to which the consumer merely added water. Orange barley water was introduced soon after.

In the 1950s, comminuted citrus drinks were introduced among the concentrates and swiftly rivaled squash in popularity. They were sold as whole fruit drinks, their flavor derived not from the juice alone but also from the natural oils extracted from the peel as the whole fruit was broken down by the process of comminution to provide the base for the drink. Thus, fruit juice became an increasingly important ingredient in soft-drink manufacture. Citrus juices were extensively used: principally orange juice from the United States, Israel, and later Brazil, among other sources, but also lemon juice, traditionally from Italy (or, more specifically, Sicily), and lime and grapefruit juices. Among the temperate fruits, apple juice provided a basis for nonalcoholic ciders.

By the 1890s, an unfermented drink made from cranberry juice was being sold on the streets of St. Petersburg, but this may have been freshly expressed and unpreserved. In England, from the 1930s forward,

concentrated black-currant juice drinks exploited the vitamin C available from indigenous sources. Fruit juices, as such, began to be packed for retail sale once ways were found of applying the principles of pasteurization to their preservation. Dr. Thomas B. Welch of New Jersey was said to have set the stage for American fruit-juice processing in this way when, as early as 1869, he began producing an unfermented sacramental wine from grape juice. But it was not until the 1930s that technical advances enabled prepacked fruit juices, including tomato juice, to be retailed on a large scale. Beginning in the 1940s, concentrated and, then, frozen juices also became available.

The citrus juices, especially orange juice, supplied a growing world market during the twentieth century, particularly once the importance of vitamin C in the diet became appreciated. Pineapple from Hawaii, and later from the Philippines, led the rest of the tropical juices in popularity. Apple, pear, and grape were significant among temperate juices consumed as such, and those fruits yielding a more pulpy juice were marketed as fruit nectars. Originally sold in glass bottles or metal cans, fruit juices are now more frequently packed in aseptic cartons, which have also been used for fruit-juice drinks in many countries of the world.

The Coming of the Colas

Amid the plethora of proprietary and patent medicines of the late nineteenth century were tonics of all sorts: These were drinks often containing phosphate and claiming to improve the nervous system and combat lassitude. As soft drinks were themselves taken for refreshment, many of these tonics, designed for a similar purpose, came to be classed under the same heading, and some of the flavors they used were also to be found in drinks designed simply to refresh.

In late Victorian Britain, kola or kaola was a popular soft drink, its essential flavor having been derived from the African kola or cola nut. The nut was known to pharmacists as a source of caffeine, and kola champagne was advertised in London as a tonic and nerve stimulant. Also available was kola tonic, the kola ingredient of which was boldly advertised as "this wonderful food," containing "more nutrient and more capacity for sustaining life than any other natural or prepared article" (Harrod's 1895 *Catalogue* 1972).

Another tonic on the market was coca wine, its stimulant properties derived from the leaves of the coca shrub, which the natives of Peru and Bolivia had long been accustomed to chew as a stimulant, and of which cocaine was a derivative. It was an interest in coca wine that led Dr. J. S. Pemberton of Atlanta, Georgia, in 1886 to combine the coca and the cola in formulating his Coca-Cola, which he, too, marketed as a brain tonic. This is not to say that these were the only flavorings. Like many speciality drinks, then as now, its formulation was a unique blend of flavors closely guarded by successive proprietors.

The word *cola* itself remained generic, and as its popularity increased, so other proprietary cola-based drinks, such as Pepsi-Cola, incorporated it in their brand names. Indeed, it became an objective of the leading brands to distance themselves from their rivals by trademark registration and, if necessary, by litigation. Pemberton and his immediate successors sold their product as a syrup for the soda fountain, which was then still the major outlet for the products of the American soft-drink industry. Only later were the proprietors somewhat grudgingly persuaded to permit others to carbonate and bottle the beverage ready-to-drink.

The licensing of other manufacturers to produce a soft drink from concentrated syrup sold to them by the owner of the brand name set a pattern for the franchising system, which came not only to dominate the twentieth-century American soft-drink industry but, eventually, to promote such brands internationally. Nor was the franchise system confined to colas. The heavy cost of transporting water-based soft drinks in glass bottles was clear from the outset of the commercial industry. There were obvious advantages in carrying a concentrated flavoring rather than the finished product over long distances, provided the proprietor could establish and enforce his standards of quality control on the local bottler so as to maintain the consistency of the consumer's drink and, thus, the product's reputation. Furthermore, a national brand could be advertised much more extensively than a local product, the cost of such advertising being recouped as part of the charge made to the bottler for the concentrate supplied. The franchise system, therefore, offered an attractive option to any owner of a drink or range of drinks, whether in America or elsewhere, looking for a way to expand.

Thus, the Canada Dry Corporation of Toronto established its ginger ale and other products in the United States during the years of prohibition and thereafter in additional markets overseas. Seven-Up, an American lemon-and-lime carbonated drink, was similarly marketed, and the promoters of speciality drinks like Dr. Pepper, a cherry soda that began as a fountain syrup, used the same means to extend their sales of ready-to-drink products at home and abroad.

Franchisers of one brand could also become franchisees of another. In Britain, for example, the old-established firm of Schweppes not only sold its products via associated companies overseas but also linked with the Coca-Cola Corporation to form the production company of Coca-Cola and Schweppes Bottlers in the United Kingdom. But the international success of the franchise system presupposed the existence of a worldwide network of local soft-drink manufacturers available to take it up.

The countries of Europe and North America that had seen the earliest growth of the carbonated soft-drink industry included those nations with the keenest interest in overseas trade. Thus, the techniques of

soft-drink manufacture spread overseas along already established trade routes. In an age of imperialism, colonies tended to import from the parent country the machinery, packaging, and many of the ingredients necessary for soft-drink manufacture until such time as they might develop indigenous industries for supplying such essentials. The supply houses of the United States, increasingly important during the nineteenth century, were similarly adept at exporting to local bottlers in countries with which other American merchants were already engaged in general trade.

The success of soft-drink manufacture in different countries depended on a variety of factors: a degree of general sophistication in the country concerned, the competition of other drinks within it, the level of income of its citizens, and their social attitudes to alcoholic and other drinks – even a hot, dry climate was known to encourage soft-drink sales!

The trading nations not only developed existing overseas markets but also fostered new ones. Japan, opened to the world in the second half of the nineteenth century, saw small-scale bottling of carbonated soft drinks by the 1890s, with British techniques, and indeed British equipment, predominating in the early years. After World War II, American franchised soft drinks came onto the Japanese market. Influenced, but by no means dominated, by international trends, Japanese soft-drink manufacturers became adept at introducing new types of drinks and packaging into an increasingly dynamic and sophisticated market.

By the time franchisers began to look overseas, there was – at least in the most promising countries for development – no lack of local bottlers available to take up the franchises. And although international brands competed with local products, they could also stimulate the growth of the local soft-drink market by promoting greater consumption. In some areas where local manufacture was less advanced, the importation of soft drinks could encourage a local industry to develop. For example, many Muslim countries of the Near and Middle East relied initially on a high proportion of imported drinks to meet the demands of a hot climate, a youthful population, and the religious restrictions on – or outright prohibition of – alcoholic drinks. Then, as wealth and demand grew, a sophisticated indigenous industry developed.

Despite the spread of international brands, patterns of soft-drink consumption still varied considerably from country to country, as idiosyncrasies of national palate and social custom determined the way in which a country's total drink market was split among competing beverages. For example, in the Russian states, herbal beverages found a substantial market, with mint, nettle, coriander, and marjoram among the flavorings used. Equally, if not more, popular in Russia was *kvass*, a low-alcohol drink made from stale bread or cereals by the incomplete fermentation of alcohol and lactic acid. It remains to be seen what effect the international brands now being fran-

chised in Russia will have on these traditional flavor preferences.

The Growing Market

Many soft drinks have been seen to owe their origins to notions of a healthy diet current at the time they first appeared, and many continued to remain popular by appealing especially to young people. But in the later twentieth century, other soft drinks were developed that reflected the desire for a healthful diet while also appealing to older people who might hitherto have thought they had outgrown the soft drinks of their youth. Low-calorie soft drinks, designed specifically for weight-conscious adults, arrived in the 1960s. These products used a new artificial sweetener, cyclamate, which blended successfully with saccharin to make a more palatable product. When, later in the decade, cyclamate was banned in many countries, saccharin-only low-calorie drinks proved less acceptable. But with the introduction of new intense sweeteners in the 1980s, the low-calorie market once again expanded.

Natural mineral waters retained their popularity over the years in many countries of continental Europe, but in the final quarter of the twentieth century, vigorous promotion, spearheaded by the Perrier brand, revived dormant markets elsewhere and developed new ones for both carbonated and still waters. Flavored, but still unsweetened, variants of such natural waters further extended their newfound popularity.

A belief in the healthful properties of natural products – and its obverse, a distaste for additives – led to the development of so-called new-age drinks, which are clear and lightly carbonated, with unusual fruit and herbal flavorings, either singly or in combination. They have no added coloring, no salt, no caffeine, and little or no added sugar. For a quite different health market, isotonic or sports drinks were produced, high in sucrose (or sometimes maltose) and designed to quickly replace the body fluids and salts excreted in vigorous exercise.

Alternative refreshment was also increasingly sought in iced tea, a water-based drink, often flavored with soft fruits and, sometimes, with herbs. In these ways the soft-drink industry has continued to innovate and expand in the 200 years of its commercial existence.

Soft Drink Packaging

A history of soft drinks – and certainly of carbonated soft drinks – could scarcely be considered complete without mention of its containers and closures, for if a carbonated soft drink is to be consumed as its manufacturer intended, it must be packed in a container capable of withstanding the pressure of the gas within and sealed with a closure that is not only effective when in place but also readily responsive to the

purchaser's efforts to open it. That these several requirements are obvious does not make them necessarily compatible!

The effective sealing of bottles presented problems from the very outset of commercial manufacture. For this purpose, bottles and jars of stoneware or of stout glass were stoppered with corks that needed to be tight or, better still, wired on to the container. Drinks subject to secondary fermentation, such as brewed ginger beer, were especially likely to burst the cork (hence ginger pop) and traditionally came to be packaged in stoneware containers.

Some early glass bottles were oval ended at the base so that they had to be stored on their sides, thus keeping the corks moist and expanded for better sealing. The bottles themselves were usually returnable when empty and were embossed with the bottler's name to promote their safe return. The inconvenience to the customer of the oval-ended bottles – they were sometimes called drunken bottles precisely because they could not stand up – led to their gradual replacement by those with a conventional flat base. These came into use around 1870 but at first were still intended to be stored on their sides.

Early manufacturers despatched their goods in strong wicker baskets, but wooden crates strengthened with wire came to be preferred, particularly once the flat-bottomed bottles arrived. But the modest price of the drinks, relative to their total weight when bottled and crated, made transportation costs an increasingly significant proportion of the whole trading operation. Although the convenience of a flat-base bottle was undoubted, its implications for the drying out of the cork soon led to improvements in bottle sealing.

Beginning in the 1870s, returnable internal screw stoppers became available, made first of hard woods and later of ebonite. Another successful invention of the time was Hiram Codd's bottle with the round glass "marble" stopper in the neck, which the pressure of carbonation kept sealed against the bottle mouth. In the last decade of the century, William Painter's crown cork came to be used for sealing the smaller returnable bottles for mixer drinks. The effectiveness of all the new seals was much improved when fully automatic bottle production superseded hand-blown glass during the first years of the twentieth century.

The internal screw stopper could be opened manually by a sufficiently strong wrist. The Codd's bottle, too, could be opened manually by depressing the marble in the neck, but wooden openers were available to supply any extra force needed. For the crown cork, a bottle opener was necessary and the closure was discarded after use. Returnable bottles and closures had also to be thoroughly washed before reuse, and the inspection of Codd's bottles and stoneware jars after cleaning could cause particular problems for the bottler.

Returnable glass bottles in sizes seldom greater than a quart dominated the industry during the first

half of the twentieth century and beyond. They usually bore paper labels, although some had their labeling information permanently embossed or fired onto the container. In the latter half of the century, packaging diversified considerably. Lighter, standardized returnable glass bottles no longer needed to bear their owners' embossed names but consequently became less distinguishable from the throwaway containers of most other bottled goods.

From the 1960s on, often in response to changes in retailing – in particular the growth of supermarkets – carbonated soft drinks were also packed in nonreturnable glass and in steel or, later, aluminum cans. The first cans required special openers, but the ring-pull end was later introduced for manual opening.

In the 1980s, soft-drink bottlers began to make blow-molded plastic containers of polyethylene terephthalate (abbreviated to PET) in sizes larger than the traditional capacities of glass bottles. Even for returnable bottles, the nonreturnable aluminum cap, rolled on to the external screw thread of the bottle neck, gained ground rapidly in the 1960s. Later, resealable closures became commonplace, particularly as bottle capacities increased.

As the second half of the century progressed, draft soft-drink equipment was more frequently introduced into catering establishments. Draft soft drinks were produced either by the premix method, whereby the drinks were carbonated and packed at the manufacturing plant and then taken to the retail outlet to be connected to the dispensing equipment on site, or by the postmix method, whereby the ingredients of the drink were loaded separately into the dispensing equipment, itself designed to mix them with water in preset proportions for dispensing as a finished, ready-to-drink beverage. By midcentury, coin-operated automatic vending machines also sold prepacked soft drinks, and there, too, in the years that followed, new versions were developed that dispensed carbonated drinks into cups by means of either the pre- or post-mix method.

Laminated cartons were developed for the packaging of fruit juices and were also used for still soft drinks. This diversification of packaging was also significant in stimulating the growth of the industry in the late twentieth century by ensuring the availability of soft drinks in an ever-increasing number of retail outlets, so that they may now be said to be among the most widely sold manufactured products in the world.

Colin Emmins

Bibliography

- Barber, R., ed. 1988. *The worlds of John Aubrey*. London.
 Barlowe, Arthur. [1589] 1970. The discovery of Virginia. In *The Tudor venturers: Selected from 'The principal navigations, voyages, traffics and discoveries of the*

- English nation made by sea or over land,' by Richard Hakluyt, ed. J. Hampden. London.*
- Bateson, M., ed. 1901. *Records of the borough of Leicester*. 4 vols. London.
- Carpenter, K. J. 1886. *The history of scurvy and vitamin C*. New York.
- Chadwick, Edwin. 1833. *The Poor Laws of London and Berkshire*. London.
- Darwin, Erasmus. 1794. *Zoonomia*. London.
- Dodd, George. 1856. *The food of London*. London.
- Fuller, T. [1662] 1965. *The history of the worthies of England*. London.
- Goold, Joseph. 1880. *Aerated waters and how to make them*. London.
- Harrod's Stores, Ltd. 1972. *Victorian shopping: Harrod's catalogue, 1895. Introduced by Alison Adburgham*. London.
- Henry, T. 1781. *A mode of impregnating water in large quantities with fixed air*. Warrington, England.
- Hicks, D., ed. 1990. *The production and packaging of non-carbonated fruit juices and fruit beverages*. New York.
- Hunt, T., ed. 1862. *Correspondence of Leigh Hunt*. 2 vols. London.
- Jekyll, G., and S. R. Jones. 1939. *Old English household life*. London.
- Kirkby, W. 1902. *The evolution of artificial mineral waters*. Manchester.
- Magellan, J. H. de. 1777. *Description of a glass apparatus for making . . . the best mineral waters of Pyrmont, Spa, Seltzer, Seydschutz, Aix-la-Chapelle. . .* London.
- Martin, M. W. 1962. *Twelve full ounces (Pepsi-Cola)*. New York.
- Mayhew, Henry. 1851. *London labour and the London poor*. 3 vols. London.
- Mitchell, A. J., ed. 1990. *Formulation and production of carbonated soft drinks*. New York.
- Nichols, S. 1994. *Vimto - the story of a soft drink*. Preston, England.
- Pendergrast, M. 1993. *For God, country and Coca-Cola*. London.
- Pereira, Jonathan. 1843. *Treatise on food and diet*. London.
- Priestley, J. 1772. *Directions for impregnating water with fixed air*. London.
- Pringle, J. 1774. *A discourse on the different kinds of air*. London.
- Riley, J. J. 1958. *A history of the American soft drink industry*. Washington, D.C.
- Rutt, J. T., ed. 1831. *Life and correspondence of Joseph Priestley*. 2 vols. London.
- Simmonds, W. H. 1905. *The practical grocer*. 4 vols. London.
- Simmons, D. A. 1983. *Schweppes - the first two hundred years*. London.
- Sulz, Charles H. 1888. *A treatise on beverages or the complete practical bottler*. New York.
- Tressler, D. K., and M. A. Joslyn. 1961. *Fruit and vegetable juice processing technology*. Westport, Conn.

which it was made a commodity of world importance by the British and Dutch East India Companies. Wars have been waged, nations punished, and fortunes made and lost because of this beverage. One reason that tea has been a very profitable article of trade is that it is a source of caffeine - a major factor in its popularity (Willson and Clifford 1992).

Although there are many kinds of herbal beverages called "teas," in this chapter the term "tea" refers only to *Camellia sinensis*. Both infusions (pouring hot water over the leaves) and decoctions (boiling the leaves in water) have been made from the plant, which has also been eaten raw, cooked, or pickled and has even been snuffed.

Botany and Production

Camellia sinensis is native to the mountainous highlands between India and China and, when untrimmed, can grow to a height of about 10 meters. One variety, *Camellia sinensis sinensis*, is native to China, whereas *Camellia sinensis assamica* comes from India (Harler 1963). A semitropical or tropical climate is necessary for raising tea; the northernmost places where it has been grown are South Carolina (in the United States) and in Asia near the Black Sea. Bushes grown at high elevations produce the best tea (Willson and Clifford 1992). Yields range from 700 to more than 1,800 kilograms (kg) per acre (Harler 1964).

Tea undergoes a long journey from bush to cup. The leaf buds, which appear in the spring, along with the leaves directly below them on the stem, are generally hand plucked by skilled workers (Pratt 1982). The first buds to appear (the "first flush") and the smallest leaves ("two leaves and a bud" is the ideal) command the highest price, whereas the larger leaves that can be cut by machines usually become part of lower-grade teas.

About 98 percent of the tea that enters the world market is black tea, and there are four steps required to ready this tea for Western tastes. The first, "withering," involves drying the leaves to the point at which they are structurally weak. This is followed by "rolling" the leaves (in the past, between the heels of the hands), which crushes their cells and blends their chemical constituents, the characteristic tea polyphenols and the enzyme polyphenol oxidase. Moreover, the twist this imparts to the leaves slows the rate at which their essence blends with hot water. "Fermentation" - the third step - actually involves oxidation by the polyphenol oxidase and is produced by aerating and heating the leaves to between 72° and 82° Fahrenheit for 1 to 2 hours, during which their flavor develops and they turn a brownish color. The final stage is called "firing" the leaves. This further heating process stops fermentation (oxidation) by deactivating the enzyme and reduces the

III.11 ☞ Tea

Tea, a drink made from the leaves and buds of the shrub *Camellia sinensis*, is the most culturally and economically significant nonalcoholic beverage in the world. Originating in China, it had spread to surrounding nations before European contact, after

leaves to a moisture content of about 5 percent (McGee 1984). Needless to say, all of these procedures require a highly developed sense of timing on the part of the tea maker.

Although black teas monopolize the international market, the teas drunk in China and Japan are mostly green (Harler 1963). In preparing the latter, the fermentation stage is eliminated, and the enzymes are destroyed by steam or pan heating before the leaves are rolled and fired. In North Africa, too, green tea is common, in this case because the classic word "fermentation," applied to the oxidation process, was misunderstood by Muslim religious and political leaders; they thought the word referred to the formation of alcohol, which their religion prohibited. Thus, this stage of tea preparation was omitted.

The size of the leaves determines the grade of a green tea: The tight, small balls of younger leaves are called "gunpowder"; "imperial" tea comes in larger and looser balls; Hyson varieties are long and looser leaves of mixed ages (McGee 1984). It might be added at this point that Oolong tea stands between green and black teas in that although it is fermented, this is only done briefly for about 30 minutes after rolling.

Grades of black teas, although they may vary, typically conform to guidelines that emerged from the British tea industry in nineteenth-century India and Ceylon (Sri Lanka). Whole leaves, which are generally thought to produce the best flavor, are classified by size and the ways in which they are rolled. For example, Flowery Orange Pekoe leaves are smaller than those of Orange Pekoe, and some of the leaves of Golden Flowery Orange Pekoe have highly prized golden tips (indicating that the tea consists of small, intact leaves), whereas in the case of Tippy Golden Flowery Orange Pekoe, all the tips are golden.

Leaves that have been broken, whether deliberately or not, also make high-quality teas that conform to some of these classifications (for example, Broken Orange Pekoe, Golden Broken Orange Pekoe, or Tippy Golden Broken Orange Pekoe). Small pieces of leaves (debris from the processing of whole and broken leaves) are called "dust," and even smaller ones are called "fannings" (Stella 1992). These go into teabags and brick tea.

Teas are also named for the regions where they grow. For example, black teas from eastern India are known collectively as Darjeeling teas; Oolongs and Lapsang Souchongs come from Taiwan, and Keemun is a black tea from northern China. Teas are sometimes scented: An extreme case is Lapsang Souchong, a large-leafed black tea scented with pinewood smoke (Goodwin 1991). More often, flavoring is done with essential oils before, during, or after firing. This is especially the case with blends – the form in which most teas are generally purchased. Constant Comment, for example, is a blend from Sri Lanka flavored with orange peel; Earl Grey, another blend, is flavored with a citrus oil (McGee 1984).

History

Origins

The true origins of tea are unknown. Wild tea leaves are still used by the tribes of Burma to prepare a beverage and a "tea salad" (made from leaves that have fermented underground for several months and are then mixed with mushrooms, oil, garlic, chilli peppers, and perhaps other ingredients). In that part of the world, tea is also chewed or sniffed as snuff, and many surmise that use of the tea plant originated there and spread to China by the Han dynasty period (206 B.C. to A.D. 221). However, the wild tea plants of nearby Assam, in India, do not produce a palatable brew, and present-day Indian tea culture is wholly the work of the British.

China

The question of whether tea came from elsewhere to Han-dynasty China is obscured by legend. One of the earliest tales would have it that (some 5,000 years ago) a mythical emperor, Shen Nung, drank boiling water into which a wild tea leaf had fallen and, consequently, was the first to taste the beverage. Other legends state that tea bushes first sprang up when a Buddhist monk cut off his eyelids to keep from falling asleep while meditating – this tale, perhaps, was inspired by the Buddhists' heavy use of tea.

Mythology, however, is not as useful to the historian in dating the Chinese use of tea as are the mentions of an herb called *t'u* in the writings of Confucius (551 to 479 B.C.), who remarked "Who can say that *t'u* is bitter? It is as sweet as the shepherd's purse" – which was another herb (Evans 1992: 14). In Chinese, the written character *t'u* later meant tea, so it is at least possible that Confucius knew of the beverage at this early date (Blofeld 1985). At any rate, there seems little question that at least some tea was drunk during the Han dynasty and that the lacquer cups known from this period (the "winged cups") are in fact the earliest teacups (Evans 1992).

Tea was probably used first for medicinal reasons, but the leaf and its lore moved from the medicinal to the artistic during the Tang dynasty (A.D. 618 to 907), a period of increasingly eclectic food and drink choices, which also saw the Chinese enjoying grape wine and butter cakes (Tropp 1982). In about 800, Lu Yu, the first known tea connoisseur, penned the oldest surviving tea manual, the *Cha Ching* or "Scripture of Tea" (Ukers 1935). This book described an elaborate ceremony of tea making, from picking the leaves, to roasting them, to serving the green tea in wide shallow bowls (Lu Yu 1974). All tea at this time was green, which kept rather poorly unless made into cakes (Blofeld 1985), and the Tang-dynasty Chinese also enjoyed spiced tea, made by adding onions, garlic, mint, orange peel, salted black soybeans, or five-spice powder (Blofeld 1985). Although despised by Lu Yu, it was a common drink, and the addition of

these flavorings may have either disguised the flavor of poor tea or allowed more bowls to be made from the same amount of leaves. Of course, such beverages also – at least to some extent – prefigured the flower teas, such as “Jasmine” and “Rose Congou” later served in North China, and which subsequently became popular outside China.

Tea ways changed as tea use and tea plantations grew more widespread during the Sung dynasty (960 to 1280). Green tea was powdered, whipped until it was the color of a “river of jade,” and served in deeper and wider bowls (Evans 1992). The use of spices and other additives was discontinued. Poets praised tea and referred to *cha tao*, the “Way of Tea” (Blofeld 1985), and the beverage was used in Chan (Zen) Buddhist ceremonies.

Brick tea was employed by the Chinese to buy horses from the central Asian tribes, and in Tibet and Mongolia, it was churned with yak butter and barley or millet into a porridge, the national staple (Kramer 1994). Centuries of Mongol rule in China were followed by the Ming dynasty (1368 to 1644), under which the tea trade widened in scope. To preserve tea for trade, some varieties were fermented into novel black teas that had not previously existed (Evans 1992). Wine pots were employed as teapots for steeping leaf tea, and by 1500, such pots were being made especially for tea (Chow and Kramer 1990). It is entirely possible that the use of a pot, enabling the brew to be poured off the leaves, was connected to the increasing importance of leaf tea, for the brew – especially in the case of black teas – can turn bitter if left in contact with the leaves too long (Chow and Kramer 1990).

Under the Ming, small handleless cups were used with a teapot, and this was the style of tea drinking exported to the West, along with the unglazed earthenware pots of Yixing (favored by tea connoisseurs) and teacups of blue and white porcelain (Chow and Kramer 1990). From the sixteenth through the eighteenth centuries in China, several styles of teapot were in common use, including some with a handle on the side and others with no handle (Kanzaki 1981). These were used for fine green tea, which must be brewed with simmering (rather than boiling) water (Arts 1988). Other types of pots included tiny personal teapots – “personal” because the tea was drunk directly from the spout (Tropp 1982).

Another Chinese development, which never reached the West, was the tea “requiring skill.” In this ritual, the finest green tea was packed into a small teapot, boiling water poured over the tea, and the first infusion discarded. The second infusion, however, was drunk as if it were a liqueur, and sometimes even a third infusion was prepared. Such an artistic tea ceremony is still practiced in Taiwan (Blofeld 1985).

Teahouses were social centers in China; they are documented as far back as the Tang dynasty and achieved great prominence from the sixteenth

through the eighteenth centuries (Anderson 1994). Patrons bought the “house tea” or had their own brewed for a small fee. Poetry readings, opera singing, or dancers were typical teahouse entertainments. At some establishments, in fact, “singsong girls” (prostitutes) offered the customer other choices in addition to tea (Blofeld 1985). Most teahouses served food: The most common fare was and remains *dimsum*, an array of snacks (Anderson 1994).

In 1644, China was taken over by the central Asian Manchu dynasty, which presided over (among other things) a new fashion of lidded cups called *chung* (Chow and Kramer 1990), in which tea could be both steeped and drunk. *Chung*, however, never became popular in Japan or the West, and even in China, the older types of pots and cups remained in use as well. The origins of many kinds of tea still enjoyed today can be traced to this, China’s last imperial dynasty. The Manchu, whose staples on the steppes had been milk and butter, also served milk in black tea, a practice which did become popular in the West (Ukers 1935).

In the early seventeenth century, the Europeans, beginning with the Dutch, sought to trade for tea with China. The Chinese, however, proved to be difficult trading partners. Not only did they claim not to need (and consequently would not accept) anything from the West except the precious metals (silver and gold) and copper but they also kept the secrets of tea growing and processing to themselves – even to the extent of boiling any tea seeds they sold to render them sterile (Goodwin 1991). Indeed, so successful were the Chinese in maintaining European ignorance of matters concerning tea that it was only in the nineteenth century (after some two and a half centuries of tea trading) that their customers learned that black and green tea came from the same plant (Ukers 1935).

The largest of these customers was the British East India Company,¹ which from the 1660s onward carried huge amounts of tea from Canton to Britain but always struggled with the balance-of-payments problem presented by the Chinese insistence on precious metals as the only exchange for tea. One solution for the company was to find products that the Chinese would accept in trade. The other was to grow its own tea, and India became the place where both of these goals were pursued.

In pursuit of the latter solution, Lord Bentinck (William Henry Cavendish), who was appointed governor of India in 1828, created the Tea Committee to foster a tea industry in that land. Several botanists were sent to China to learn about tea cultivation and to recruit Chinese growers to start tea plantations in India. One of these botanists, J. C. Gordon, collected more than 80,000 seeds and sent them to Calcutta, whereupon the British discovered that all varieties of tea came from the same plant (Willson and Clifford 1992).

But it was Robert Fortune – entering China in 1848 as an agent of the Tea Committee and for three

years roving about the country's tea regions disguised as a Chinese merchant – who finally unlocked the many mysteries involved in actually producing tea (Beauthéac 1992). Thereafter, Chinese tea was planted in India and, a bit later (after disease had wiped out the island's coffee trees), in Ceylon (Sri Lanka) as well. Tea bushes native to India were subjected to experimentation, hybrids tinkered with, and, ultimately, plants well suited to India's climate were developed. The great tea plantations of British India were born.

Nonetheless, because China continued to supply most of the world's tea, the problem of its insistence on bullion or copper as payment for that tea also continued. Actually, the British had already found one product – cotton from Bengal – that interested the Chinese. Then they hit upon another. Poppies, imported from Turkey, grew well in India, and the British East India Company entered the opium business. No matter that the Chinese government refused to permit the importation of the addictive powder. Intermediaries soon stimulated a lively demand for it across all strata of Chinese society, and as the government continued to object, the British went to war, their gunboats easily overwhelming the Chinese coastal defenses. The Treaty of Nanking (1842) not only forced opium upon the Chinese but also forced open four Chinese ports (in addition to Canton) to European trade. Hong Kong was ceded to the British, and British consuls were admitted to all treaty ports.

The decline of China from its former imperial grandeur continued throughout the remainder of the nineteenth century as foreigners increased their ascendancy, and in the twentieth century, the old China disappeared in the tumult of war and revolution. During these years, a great deal of the ancient art and culture surrounding tea was also destroyed. But it remained (and remains) the national beverage. Throughout the day, people carry with them lidded tea mugs, which, like the *chung*, can be used both for infusion and for drinking (Chow and Kramer 1990). Glass mugs or cups are aesthetically pleasing, because they allow the drinker to view the unfolding of the tea leaves as the tea brews (Chow and Kramer 1990); in fact, there are teas which are specially bred and rolled for this purpose. Teahouses in the new China include opera, television viewing, and even “karaoke” singing as entertainment (Anderson 1994).

Japan

Japan first obtained tea, along with many other cultural practices, from China. Tradition has it that the beverage was carried to Japan by the Buddhist monk Eisai; drinking tea before meditation was for the Buddhists a practical way to keep awake, and early Zen monks drank tea long before it was known elsewhere in the islands. To make it, cakes of tea were powdered, and water was boiled in a kettle and ladled into a bowl. Following this step, the tea was whipped with

a whisk – itself made by splitting bamboo fibers and then bending them with steam into a double cone of curved spines. The early monastic “four-headed tea” was a rite in which a man served tea to guests by whipping individual bowlfuls (Hayashiya 1979).

In Japan, tea service evolved into an elaborate ceremony, continuing the Chinese Sung-dynasty use of whipped green tea, which as employed in the Japanese tea ritual still resembles jade in its deep, clear green color (Chow and Kramer 1990). When the country was unified in the sixteenth century after a long era of war among feudal lords, the cultivation and use of green tea became widespread, and the tea ceremony, which had served as an opportunity for ostentation by the nobility, was remade under the guidance of art collector and tea-master Sen Rikyū, who created the “tea of quiet taste” – a ceremony performed in a small, low hut and passed on to future generations as a tradition. Different kinds of sweets, flowers, incense, stylized conversation, calligraphy, bowls, and other utensils became associated with different “schools” of the tea ceremony, as well as months of the year, places, and so forth. Years of instruction became necessary to master the way of tea; Rikyū's sons and grandsons founded several tea schools, which carry on old traditions to this day (Sen 1979).

During the seventeenth century, leaf tea was introduced from China, and the use of teapots became common. For ordinary tea, a teapot with a spout on the front was used. For fine tea, a hot-water pot (*kyusu*) – more convenient than the older kettle – was employed (Kanzaki 1981). Cups of porcelain in the style of the Chinese Ming dynasty became common alongside the *kyusu*. Although the tea ceremony of Rikyū used powdered tea, another ceremony involving fine-leaf tea also developed, and the *kyusu* employed were often beautiful works of art (Arts 1988).

In 1854, the isolation of Japan came to an end with the Shogun's agreement to allow foreign ships to put into Japanese ports, and in the initial rush to modernize, the Japanese tea ceremony was lost. But at the start of the twentieth century, as Japanese culture underwent a revival, the ceremony became popular once again, a popularity that continues into the present. Tea also figures in popular culture. It is used regularly in homes and restaurants and can even be bought from vending machines. Green tea remains the most popular for daily use, although other kinds, including black tea and herb tea, are also drunk. In 1992, Japan exported a mere 290 tons of tea while importing 160,367 tons (FAO 1993).

India

Although a Dutch seafarer wrote of tea being eaten as well as drunk in India in 1598, accounts of earlier Indian history do not mention the use of tea or its cultivation (Pettigrew 1997). Milk and buttermilk, produced by the country's millions of sacred cattle, were the preferred beverages in India (Tannahill 1988).

The tea cultivation begun there in the nineteenth century by the British, however, has accelerated to the point that today India is listed as the world's leading producer, its 715,000 tons well ahead of China's 540,000 tons, and of course, the teas of Assam, Ceylon (from the island nation known as Sri Lanka), and Darjeeling are world famous. However, because Indians average half a cup daily on a per capita basis, fully 70 percent of India's immense crop is consumed locally (Goodwin 1991; FAO 1993). Tea in India is generally spiced and served with milk, thus incorporating two other prominent Indian products. As a cup of tea with sugar and milk may contain up to 40 calories, this is also a source of quick energy (Harler 1964).

In general, even though India leads the world in tea technology, the methods employed to harvest the crop vary with the type of tea and terrain. Fine-leaf tea is hand plucked, and hand shears are used on mountain slopes and in other areas where tractor-mounted machines cannot go. A skilled worker using hand shears can harvest between 60 and 100 kg of tea per day, whereas machines cut between 1,000 and 2,000 kg. The latter, however, are usually applied to low-grade teas that often go into teabags. The tea "fluff" and waste from processing is used to produce caffeine for soft drinks and medicine (Harler 1963).

Russia

Tea from China reached Tsar Alexis of Russia in 1618, and the Russians quickly adopted the beverage. They adapted the Mongolian "firepot" to their own purposes, creating the charcoal-fueled samovar, which boiled water in its tank and furnished the heat to make tea essence in a small pot atop the device. The essence was diluted with hot water from the tank whenever a cup of tea was desired. Samovars (now electrically heated) are still common, especially in offices and at parties and gatherings where much tea must be made at once. Lump sugar is used for sweetening, and Russians often bite a sugar lump in between sips of tea (Schapira 1982).

Tea is served in glasses with metal holders in public places and in china cups at home. Apples or cranberries (instead of costly lemons) are sometimes added along with sugar, and tea is usually drunk with milk. Pastries and sweets often accompany the beverage. The low price of tea led to its great popularity in Russia, and the tea plantations of Georgia (now a separate nation) are the most northerly in the world. The Russians, however, consider Georgia-grown tea to be inferior (Schapira 1982), and some 325,000 tons of tea were imported by the countries of the former Soviet Union in 1992 (FAO 1993).

Continental Europe

Tea first became popular in the Netherlands in the early seventeenth century. Served in the afternoon – in what would become the British style – it was mixed with sugar and, sometimes, saffron. As the cus-

tom of afternoon tea in the home developed, hostesses set aside special rooms, furnished with paintings, tables and chairs (Brochard 1992), and large quantities of tea were reportedly consumed: Montesquieu, for example, saw a Dutch woman drink 30 small cups at one sitting (Brochard 1992). In the Dutch colony of New Amsterdam in North America, water for tea was hawked through the streets, and prior to the dominance of the British East India Company, the Netherlands exported much tea to Britain (Israel 1987).

Tea cultivation in Indonesia, which began under Dutch supervision in the mid-nineteenth century (Ukers 1935), specialized in the production of black teas often used for blends. Tea is still popular in the Netherlands today; in 1992, the United Provinces imported 377,803 tons of tea and reexported 201,306 tons, leaving 176,497 tons for local use (FAO 1993).

Germany and France have historically consumed little tea, although the porcelain works of Meissen in Germany have produced excellent tea services. Some tea is consumed in the Dutch or British fashion in both nations, but most people there prefer coffee.

Britain

Despite the British fondness for tea, it does not have a long history in the islands. The first mention of "chaw" in a sea captain's letter dates from 1597, and not until 1658 was tea first announced for sale at a London coffeehouse (Ukers 1935). From that point forward, however, sugar from British-controlled islands of the Caribbean became plentiful enough to make the new beverage palatable to Englishmen's tastes. But leather or wooden cups – used for ale – were unsuitable for hot tea, and consequently Ming-style pots and teacups were brought to Britain along with the new beverage.

Expensive at first, tea became more widespread as duties fell. By the 1710s, afternoon tea had become an important convivial occasion for women, who were discouraged from drinking alcohol socially because (according to the new bourgeois code of morals) they could not remain "ladylike" while intoxicated. In addition, eighteenth-century Britain saw the rise of "tea gardens," where tea, snacks, and entertainment could be enjoyed by the bourgeois and upper classes. Sidney Mintz (1985) has commented on the "tea complex" of tea and sugary foods (such as jam and pastries) that came to be popular among the British, doubtless in no small part because of the stimulating effects of caffeine and sugar. According to Benjamin Franklin, eighteenth-century Britons and Anglo-Americans drank 10 pounds of tea per head yearly – which translated into 2,000 cups, or about 5½ cups a day (Pratt 1982).

Such demand invariably led to adulteration in order to color the tea and add weight to shipments. Ash leaves, licorice, iron filings, Prussian blue, sheep dung, and numerous other materials could be found in tea

packages, and discarded tea leaves were saved from garbage heaps and washed for resale. Even the tea-loving Chinese happily dyed green tea a brighter green to satisfy British customers who disliked the pale color of the “real thing” (Goodwin 1991). It is interesting to note that some popular scented teas are the products of such adulteration: In the case of Earl Grey, for example, buyers came to demand the scent and flavor produced by the addition of oil of bergamot.

Chinese porcelain was painted with designs made especially for English buyers (such as the famous “willow pattern”), and cups began to be made with handles to enable the drinker to cope with tea hot enough to dissolve sugar – unlike the Chinese green tea, which was brewed at a lower temperature (Pratt 1982). Carried as ballast in sailing ships (one ship might carry as much as 20 tons of porcelain), this “china” was very inexpensive (Atterbury 1982; Hobhouse 1986), but attempts were nonetheless made to copy Chinese porcelain. Such efforts eventually met with success, and by 1742 porcelain was being made in Europe.

British tea tariffs, which played such an important part in bringing about the American Revolution, were repealed the year after American independence was achieved (Goodwin 1991), and the adulteration of tea ceased as it was no longer profitable (Hobhouse 1986). In 1786, the East India Company sold a total of 2.4 million pounds of black tea and 1.15 million of green tea (Ukers 1935). Even in those days, green tea was already losing ground to black, but it was still relatively more popular than it would later become with the massive influx of black teas from India and Africa. Broken grades also sold very well in Britain because of their lower prices (Harler 1963).

The Victorian period found tea perhaps at its zenith, with even the homeless of Whitechapel partaking – a victim of Jack the Ripper was carrying tea on her person, along with sugar and a spoon, at the time of her murder (Tully 1997). The temperance movement was greatly aided by affordable tea, as millions were converted to “tee-totaling,” or abstinence from alcohol. The British took tea in the morning, at lunch, at “low tea” (afternoon tea, served in the “low” part of the afternoon), and at “high tea” (or “meat tea”), an early supper served at 6 o’clock (Israel 1987). Strongly scented teas, such as Earl Grey, Jasmine, and Lapsang Souchong, were extremely popular, perhaps because the Victorian regimen of meats and sweets demanded strongly flavored teas to match. The rituals of tea became more complex, and from the drawing rooms (the “withdrawing rooms” to which ladies withdrew while gentlemen remained at the table) of London to the South African bush, the British cherished tea, the “cups that cheer.”

In the twentieth century, British workmen continued to break daily for tea, made and served collectively for groups of workers or purchased from nearby shops. In the 1980s and 1990s, the British

each consumed some 3 to 4 kg of tea annually. In 1992, the nation as a whole imported 503,350 tons of tea and reexported 122,476 tons, leaving 380,874 tons for local consumption (FAO 1993). The rising popularity of coffee and soft drinks have somewhat diminished tea’s popularity, but it remains a staple British beverage, and tearooms flourish throughout much of the English-speaking world, although not in the United States.

The United States

Tea in America has a history going back to the colonial era, when it was used by both Chesapeake planters and Massachusetts merchants and was especially popular in Philadelphia (Roth 1988). As in England during this period, most tea was green and often mixed with sugar, fruits, and liquors to make an alcoholic “punch.”

The legendary Boston and Charleston “tea parties” sprang from the British tax on tea, which, unlike other taxes, had not been repealed in the face of colonial protest. The result was a series of civil disorders in which shiploads of tea were destroyed (in Boston) or stolen (in Charleston). In retrospect, these acts were clearly precursors of the Revolution. Contrary to legend, however, tea drinking did not cease because of hostility to Britain, and George Washington continued to breakfast on three bowls of tea. The difference was that this tea no longer reached the United States through British channels, but rather in American ships, the first shipment arriving in 1784 aboard the *Empress of China* (Ukers 1935). But it was the case that, a few decades later, an increase in the availability of cheap, Brazilian, slave-produced coffee brought a gradual decline of tea consumption in favor of coffee.

Two great changes in American tea drinking came about in the early twentieth century. One occurred in 1908, when tea merchant Thomas Sullivan, in order to reduce shipping weight, began to package tea samples in silk bags instead of miniature tins. Some of his customers brewed the tea without taking it out of the bags and requested more tea packaged in this way; Sullivan obliged, and teabags were created. Today, in America, most tea is brewed from teabags (Schapira 1982).

The other innovation was iced tea, supposedly invented at the 1904 World’s Fair by Richard Blechynden, an English tea concessionaire who had been sent to the United States to promote black tea. But in the sweltering heat of St. Louis, he was unable to sell hot tea or even give it away. In frustration, he poured the tea over ice and began to serve it cold. Because there is much evidence of iced tea before this event, Blechynden gets more credit for his “invention” than he deserves, but it is probably fair to say that he brought iced tea to the world’s attention (Pratt 1982; Israel 1987). Indeed, from this point forward in the United States, the sale of black tea (which makes better iced tea) began to edge out

that of green, suggesting that Blechynden gave his employers their money's worth.

Still another development, which many find to be no improvement at all, is the invention of "instant" tea, which is an evaporated powder similar to instant coffee (Willson and Clifford 1992). Instant tea constitutes a significant amount of the tea consumed in the United States, but – as it is not made by brewing – it could be said that it is not really tea at all (Schapira 1982). In the early 1990s, the nation imported 305,017 tons of tea and exported 85,363 tons. Included in the latter was a small amount of tea grown in the United States (FAO 1993).

Until at least very recently, tea cultivation in the United States has never been a profitable enterprise because of climate, although, since the nineteenth century, individual tea bushes have been tended in gardens of the South. Between 1848 and 1852, the operators of a plantation in Greenville, South Carolina, tried to grow tea, and in 1880, John Jackson made an attempt in Summerville, South Carolina. Tea was also grown in several places nearby, but all these efforts were eventually abandoned (Ukers 1935).

In the 1980s, however, Mack Fleming, an engineer for the Lipton Tea Company, purchased the Summerville land and, with the help of Canadian tea taster Bill Hall, began experimentally processing leaves from the tea bushes that still remained. Fleming and Hall transferred tea bushes to Wadmalaw Island near Charleston and invented a machine to shear off the tea buds, literally mowing the tea like a lawn. The resulting product, marketed in both bagged and loose forms as American Classic Tea, is the only tea commercially grown in America. Hawaii, California, and Florida also have climates appropriate for tea growing, but labor costs for harvest are prohibitive.

Other Lands

Brazil and Argentina in South America, and Turkey and Iran in the Near East, produce undistinguished teas – mostly for teabags – as do Tanzania and Malawi in Africa. In Kenya, however, which now ranks fourth among the world's tea-producing nations, some teas are emerging as noteworthy. For the most part, these black teas are grown on plantations on the high plateaus that are understandably similar to those of India, as they were founded by British planters who left that country following independence (Willson and Clifford 1992). In addition, South Africa produces an excellent tea from plants that originated in the Darjeeling area of India; in South Africa, as in Kenya, tea plantations are located on the high plateaus.

Tea and Health

In modern times, tea is one of the world's least expensive beverages, which helps to account for the fact that, after water, it is also the most commonly used beverage. Because it is made with boiled – and there-

fore sterile – water, tea is safe to consume in areas where water quality may be less than satisfactory.

This is doubtless one of the major reasons that, over the millennia, tea has maintained a reputation for contributing to good health (Weisburger 1997). Buddhist monks in China and Japan, who were physicians as well as tea-masters, used tea to help people through their illnesses. Even today, tea is frequently employed in ancillary therapy for individuals who suffer from various infections, colds, or chronic diseases (Yamanashi 1991). Tea is a diuretic and produces both warmth (if it is served hot) and coolness (because it promotes evaporation of water from the skin) in the body; moreover, the steam of hot tea moisturizes the nose, mouth, and throat. The caffeine in tea is also a "pick-me-up."

Perhaps because tea was first used in China and Japan, historically most research on its health benefits was conducted by Eastern research groups. These groups have held conferences to report on findings that have to do largely with disease prevention and, to a lesser extent, with therapy (Yamanashi 1991). But even though tea became a major beverage in some of Europe and all of the English-speaking world, relatively little research of this nature was done in the West until recently.

Beginning in about 1980, however, a number of laboratories and groups in the United States and Canada – and a few in Europe – have shown an interest in researching relationships between tea and human health. Their findings were summarized at the First International Conference on Tea and Health, held in New York in 1991 (Symposium 1992). Included were reports suggesting that people who drink tea regularly have a lessened risk of coronary heart disease. Moreover, some investigators also noted that tea has a beneficial effect in warding off cancers of the stomach, esophagus, and lung (Katiyar and Mukhtar 1996; Weisburger 1996; Yang et al. 1997), and new research suggests that the caffeine in tea may play a role in such cancer prevention. Black tea and green tea had very similar (if not identical) preventive effects, and a few findings showed delayed growth and even regression of cancers in laboratory animals.

Stimulated by such findings, considerably more research has since been conducted in Europe, North America, and Asia. One interesting area of investigation has to do with the incidence of lung cancer in Japanese men, which is lower than that in their American counterparts, even though the Japanese smoke many more cigarettes on a per capita basis. It has been hypothesized that tea drinking by Japanese men is the key factor.

The chemical basis for the preventive benefits of tea seems to be specific polyphenols in both green and black teas that can act as powerful antioxidants. It is suspected that the risk of heart disease and many types of cancer is lowered by the raised antioxidant level in the bodies of those who drink tea and, con-

versely, that the conditions promoting these diseases involve abnormally low levels of oxidation. Tea polyphenols also modify intestinal bacterial flora, leading to a decrease in the levels of undesirable bacteria and contributing to the maintenance and increase of desirable bacteria. In fact, tea is the culture medium for the "tea fungus" called *kombucha*, actually a symbiotic bacterial culture used in alternative medicine. Tea can have a specific antibacterial action and, perhaps, a limited but definite effect against specific viruses (Hara 1994). Moreover, recent results suggest that Alzheimer's disease may occur less frequently in aging people who are regular tea drinkers.

One kilogram of tea makes about 440 cups, whereas the same amount of coffee makes about 88 (Harler 1964), which helps to explain both the relative cheapness of tea as a beverage and the higher caffeine content of coffee. The caffeine in tea was once called "theine," but is chemically identical to the caffeine in coffee. Depending on the type of tea, a 150-milliliter cup may contain 30 to 50 milligrams of caffeine, only about one-third of the caffeine found in a cup of coffee (Harler 1963).

Nonetheless, this is sufficient to account for its reportedly pleasant, slightly stimulating effect on mental function, and is doubtless the reason for tea's popularity as a morning beverage. However, the amount of caffeine in tea is low enough that an overdose is unlikely, whereas overdose symptoms ranging from headaches to severe gastrointestinal distress do occur in coffee drinkers (Willson and Clifford 1992). Tea is also available decaffeinated; to achieve this, solid carbon-dioxide extraction is employed to remove caffeine selectively without removing the desirable polyphenols. Thus, decaffeinated tea provides the taste and most of the health benefits of its caffeinated counterpart for caffeine-sensitive tea drinkers and can be consumed in the evening without fear of sleep deprivation.

Ordinary tea, however, can be made as free of caffeine as decaffeinated tea. One way to do this is to brew a first infusion, discard it, and brew a second from the same tea. Caffeine is extracted from tea in 1 to 2 minutes, and so a second infusion will have a weaker taste but no caffeine. Another means is to make "sun tea" by placing cold water with teabags in it in direct sunlight, usually for several hours. The result is a tea virtually free of caffeine, normally served iced. However, it is not known whether the health-promoting antioxidant tea polyphenols remain intact after being subjected to the sun's radiation.

How to Make Tea

Starting with Lu Yu and Sen Rikyû, tea-masters throughout the ages have emphasized both the ease of making good tea and the philosophical significance of the beverage. The following is offered as a basic modern outline of how good tea should be treated.

Start with fine tea, whether whole or bagged, and clean, cold water. Avoid water that has a taste of its own. For green tea, bring the water to a simmer; for black or Oolong, it must be at a rolling boil. Warm the vessel in which the tea will brew by pouring boiling water into and out of it, and add about 3 grams of tea per cup to be brewed. Pour boiling water over the leaves and allow to steep for 3 to 5 minutes, depending on the type of tea. Finally, remove the leaves or teabag from the vessel and pour the tea, adding sugar (or other sweetener), milk, or lemon, as desired. On a cold night, tea with rum, brandy, or bourbon whiskey makes a warming and thus a pleasant beverage.

*John H. Weisburger
James Comer*

Research in Weisburger's laboratory is supported by grants and contracts from the Tea Trade Health Research Association, Toronto, Canada; the National Cancer Institute, Bethesda, Maryland; the American Cancer Society, Atlanta, Georgia; and gifts from Texaco, Inc., Beacon, New York, and the Friends Against Cancer Team.

Note

1. This company of many names was originally founded as two separate companies: (1) the Governor and Company of Merchants Trading into the East Indies, and (2) the English Company Trading to the East Indies. In 1708, the two merged to become the United Company of Merchants Trading to the East Indies, which was more succinctly called the Honourable East India Company and more familiarly referred to as the John Company (Harler 1964).

Bibliography

- Anderson, E. N. 1994. Food. In *Handbook of Chinese popular culture*, ed. Patrick Murphy and Wu Dingbo, 35-54. London.
- Anderson, Jennifer. 1991. *An introduction to Japanese tea ritual*. Albany, N.Y.
- Arts, P. L. W. 1988. *Tetsubin: A Japanese waterkettle*. Groningen, Netherlands.
- Atterbury, Paul, ed. 1982. *The history of porcelain*. New York.
- Beauthéac, Nadine. 1992. Tea barons. In *The book of tea*, trans. Deke Dusinberre, 57-99. Paris.
- Blofeld, John. 1985. *The Chinese art of tea*. Boston.
- Brochard, Gilles. 1992. Time for tea. In *The book of tea*, trans. Deke Dusinberre, 101-95. Paris.
- Burwell, Letitia. 1895. *A girl's life in Virginia*. New York.
- Castile, Rand. 1971. *The way of tea*. New York.
- Chow, Kit, and Ione Kramer. 1990. *All the tea in China*. San Francisco.
- Deetz, James. 1977. *In small things forgotten*. Garden City, N.J.
- Evans, John C. 1992. *Tea in China: The history of China's national drink*. New York.

- FAO (Food and Agriculture Organization of the United Nations). 1993. *FAO trade yearbook, 1992*. Rome.
- Farb, Peter, and George Armelagos. 1980. *Consuming passions*. Boston.
- Goodwin, Jason. 1991. *The gunpowder gardens*. New York.
- Hara, Y. 1994. Prophylactic functions of tea polyphenols. In *Food phytochemicals for cancer prevention*. II. *Teas, spices, and herbs*, ed. C. T. Ho, T. Osawa, M. T. Huange, and T. R. Rosen, 34-50. Washington, D.C.
- Harler, Campbell. 1963. *Tea manufacture*. Oxford.
1964. *The culture and marketing of tea*. London.
- Hayashiya, Seizo. 1979. *Chanoyu: Japanese tea ceremony*. New York.
- Hobhouse, Henry. 1986. *Seeds of change*. New York.
- Hudson, Charles, ed. 1979. *Black drink*. Athens, Ga.
- Israel, Andrea. 1987. *Taking tea*. New York.
- Kakuzo, Okakura. 1989. *The book of tea*. New York.
- Kanzaki, Noritake. 1981. *Japanese teapots*. Tokyo.
- Katiyar, S. K., and H. Mukhtar. 1996. Tea consumption and cancer. *World Review of Nutrition and Dietetics* 79: 154-84.
- Kramer, Ione. 1994. Tea drinking and its culture. In *Handbook of Chinese popular culture*, ed. Patrick Murphy and Wu Dingbo, 55-76. London.
- Lender, Mark Edward. 1987. *Drinking in America*. New York.
- Levenstein, Harvey. 1988. *Revolution at the table*. New York.
- Lu Yu. 1974. *The classic of tea*, trans. Francis Ross Carpenter. New York.
- McGee, Harold. 1984. *On food and cooking*. New York.
- Mintz, Sidney. 1985. *Sweetness and power*. New York.
- Pettigrew, Jane. 1997. *The tea companion*. New York.
- Post, Elizabeth. 1965. *Etiquette*. New York.
- Pratt, James Norwood. 1982. *The tea lover's treasury*. San Francisco.
- Root, Waverly. 1976. *Eating in America*. New York.
- Roth, Rodris. 1988. Tea-drinking in eighteenth-century America: Its etiquette and equipage. In *Material life in America 1600-1860*, ed. Robert Blair St. George, 439-62. Boston.
- Schafer, Charles, and Violet Schafer. 1975. *Teacraft: A treasury of romance, rituals, and recipes*. San Francisco.
- Schalleck, Jamie. 1971. *Tea*. New York.
- Schapira, Joel. 1982. *The book of coffee and tea: A guide to the appreciation of fine coffees, teas, and herbal beverages*. New York.
- Schivelbusch, Wolfgang. 1992. *Tastes of paradise*. New York.
- Sen, Soshitsu. 1979. *Chado: The Japanese way of tea*. New York.
- Smith, Hedrick. 1976. *The Russians*. New York.
- Stella, Alain. 1992. Tea gardens. In *The book of tea*, trans. Deke Dusenberre, 21-55. Paris.
- Symposium. 1992. Physiological and pharmacological effects of *Camellia sinensis* (tea): First International Symposium. *Prev. Med.* 21: 329-91, 503-53.
- Tannahill, Reay. 1988. *Food in history*. New York.
- Taylor, Joe Gray. 1982. *Eating, drinking, and visiting in the South*. Baton Rouge, La.
- Tropp, Barbara. 1982. *The modern art of Chinese cooking*. New York.
- Tully, James. 1997. *Prisoner 1167: The madman who was Jack the Ripper*. New York.
- Ukers, W. H. 1935. *All about tea*. New York.
- Weisburger, John H. 1996. Tea antioxidants and health. In *Handbook of antioxidants*, ed. E. Cadenas and L. Packer, 469-86. New York.
- Weisburger, John H. 1997. Tea and health: A historical perspective. *Cancer Letters* 114: 315-17.

Willson, K. C., and M. N. Clifford, eds. 1992. *Tea: Cultivation to consumption*. London.

Yamanashi, T., ed. 1991. *Proceedings, International Symposium on Tea Science*. Shizuoka, Japan.

Yang, C. S., M. J. Lee, L. Chen, and G. Y. Yang. 1997. Polyphenols as inhibitors of carcinogenesis. *Environmental Health Perspectives* 4: 971-6.

III.12 Water

The ingestion of water in some form is widely recognized as essential for human life. But we usually do not consider water as food because it does not contain any of those substances we regard as nutriment. Yet if its status as a foodstuff remains ambiguous, it is far less so than it has been through much of human history. Water (or more properly "waters," for it is only in the last two centuries that it can really have been viewed as a singular substance) has been considered as food, a solvent for food, a pharmaceutical substance, a lethal substance, a characteristic physiological state, and a spiritual or quasi-spiritual entity.

This chapter raises questions about what sort of substance water has been conceived to be and what nutritional role it has been held to have. Moreover it also explores what we know of the history of the kinds of waters that were viewed as suitable to drink - with regard to their origins, the means used to determine their potability, and their preparation or purification. It also has a little to say about historical knowledge of drinking-water habits (i.e., how much water did people drink at different times and situations?) and water consumption as a means of disease transmission.

What Water Is

Modern notions of water as a compound chemical substance more or less laden with dissolved or suspended minerals, gases, microorganisms, or organic detritus have been held at best for only the last two centuries. Even earlier ideas of water as one of the four (or five) elements will mislead us, for in many such schemes elements were less fundamental substances than dynamic principles (e.g., in the case of water, the dynamic tendency is to wet things, cool them, and dissolve them) or generic labels for regular combinations of qualities. In one strand of Aristotelianism, for example, water can be understood as matter possessing the qualities of being cold and wet; thus whenever one finds those characteristics, one is coming across something more or less watery. It may even be inappropriate to think of water as wholly a natural substance; as we shall see, springs and wells (if not necessarily the water from them) often held sacred status. The primacy of water in the symbolism of many of the world's religions as a medium of disso-

lution and rebirth invites us to recognize water as numinous in a way that most other foodstuffs are not (Eliade 1958; Bachelard 1983).

At the very least, it is clear that through much of Western history “water” referred to a class of substances. “Waters” varied enormously both in terms of origin (rainfall, snowmelt, dew, and pond, spring, and river water were seen to be significantly different) and from place to place, just as climate and other geographical characteristics – vegetation, soil, topography – vary. Whereas for most of us the modern taxonomy of water quality includes only two classes (pure and impure), in the past subtle and complicated characterizations were nearly universal, especially with regard to water from springs, a matter that fascinated many writers. Indeed, the uniqueness of a water is a key attribute of place, and waters are linked to places in much the same way in which we now associate the characteristics of wines with their places of origin. This idea is evident in the title of the famous Hippocratic treatise “On Airs, Waters, and Places”; it is a central theme in the treatments of post-Hippocratic classical authors like Pliny the Elder and Marcus Pollio Vitruvius, and much the same sensibility seems apparent in Celtic, Teutonic, and Chinese perspectives.

Water varied from place to place in many ways, but one can generalize about the kinds of qualities that interested classical authors. Many did mention taste, and they usually held that the less taste the water had, the better. Taste, in turn, was associated with a host of other qualities that linked the immediate sensory experience of drinking a water with the effects of its continued consumption on health and constitution. Other key factors were coldness, lightness, and heaviness.

Coldness did not necessarily refer to temperature, which in any case could not be measured except subjectively and indirectly. In Chinese cosmology coldness and hotness were part of the universal system of polarities, applied to foodstuffs as well as to many other substances and activities. Water was by definition cool; steeping or cooking things in it (even boiling them) was accordingly a cooling process (Simoons 1991: 24). For the mechanical philosophers of early eighteenth-century Europe, “cold” and “hot” had become terms of chemical composition: Sulphurous water was “hot,” that containing niter or alum was “cold” (Chambers 1741).

Lightness appears usually as a subjective quality akin to ease of digestibility. Pliny, in an unusual outburst of skepticism, observed that unfortunately “this lightness of water can be discovered with difficulty except by sensation, as the kinds of water differ practically nothing in weight” (Pliny 1963: book 31, chap. 21). But occasionally lightness was a parameter that could be objectively measured: Equal areas of tissue were wetted in different waters and their weights compared to determine which water was the lighter, and hence the better to drink (Lorcin 1985: 262).

The Classes of Waters

In general, desirable or undesirable properties were associated with the source of the water one obtained, and there was general agreement on the ranking of these sources. Rainwater was usually held to be the best water even though it was also regarded as the quickest to become putrid (though this in itself was not problematic, since the water might still be used after it had finished putrefying). Though some wealthy Romans made much of its virtues, water from melted snow or ice was generally viewed as harmful, possibly because it was associated with goiter, but more likely because, being deaerated, it tasted flat and led to a heaviness in the stomach (Burton 1868: 241; Vitruvius 1960: 239; Pliny 1963: book 31, chap. 21; Diderot and D'Alembert 1969; Soyer 1977: 296; Simoons 1991: 491–2).

Most deemed water from mountain streams (particularly on north-facing slopes) better than water from wells or streams in hot plains because it was held that the heat of the sun was likely to drive off the lighter or best parts of the waters (though Burton, summarizing classical authors, insisted that waters from tropical places were “frequently purer than ours in the north, more subtle [*sic*], thin, and lighter” [Burton 1868: 241]). Running waters were to be preferred to stagnant waters; waters stored in cisterns were undesirable because they accumulated “slime or disgusting insects” (Pliny 1963: book 31, chap. 21). But none of these generalizations obviated the need to characterize each individual source of water because, as Pliny noted, “the taste of rivers is usually variable, owing to the great difference in river beds. For waters vary with the land over which they flow and with the juices of the plants they wash” (1963: book 31, chap. 32).

Water and Health

Such descriptions clearly indicate that classical authors were much concerned with the effects of waters on health. One can understand their views in terms of a four-part scheme for classifying waters. Some waters are seen as positively beneficial to health, medicaments to be taken to cure various maladies. Others are viewed as good, “sweet” (the term has long been used to describe waters) waters, of acceptable taste and suitable for dietetic use. Still others are regarded as having undesirable qualities as a beverage. Accordingly, they are to be used sparingly, only after treatment, or with compensatory foodstuffs. Finally, authors recognize some waters as pathogenic in some sense, even lethal.

Beyond taste and health effects, waters (particularly spring waters) were characterized in terms of a host of bizarre properties they were believed to have. Pliny tells us (as does Vitruvius) of springs that turn black-wooled sheep into white-wooled sheep, that cause women to conceive, that endow those who drink of them with beautiful singing voices, that

petrify whatever is dipped into them, that inebriate those who drink from them or, alternatively, make those who drink of them abstemious (Pliny 1963: book 31, chaps. 3–17). To this one might add the Levitical “bitter waters of jealousy,” which possessed the property of identifying adulteresses.

Sometimes authors were explicit in attributing the properties of springs to a concept of chemical admixture: The water had the properties it had owing to what had happened to it, such as the kinds of mineral substances it had encountered underground. Yet the earth was seen as in some sense alive, and we would accordingly be unwarranted in assuming that a modern concept of solution is implied (Eliade 1978). Waters were earth’s vital fluids. The Roman architect Vitruvius noted, for example:

[T]he human body, which consists in part of the earthy, contains many kinds of juices, such as blood, milk, sweat, urine, and tears. If all this variation of flavors is found in a small portion of the earthy, we should not be surprised to find in the great earth itself countless varieties of juices, through the veins of which the water runs, and becomes saturated with them before reaching the outlets of springs. In this way, different varieties of springs or peculiar kinds are produced, on account of diversity of situation, characteristics of country, and dissimilar properties of soils.” (Vitruvius 1960: 241–2)

It is probably right to see this linkage of macrocosm and microcosm as something more than analogical; such linkages would remain a part of popular understanding even after the rise of a mechanistic cosmology in the seventeenth century.

The properties of waters might also be understood as manifesting the spirits or resident divinities of springs, because many springs and rivers were thought of as home to (or the embodiment of) a divinity. Such views were held in many premodern cultures, although perhaps best known are the 30,000 nymphs associated by Hesiod with springs in Greece (their brothers were the rivers) (Hesiod 1953: 337–82; Moser 1990; Tölle-Kastenbein 1990). In many, particularly rural, places in France, Britain, Germany, and elsewhere, worship of such divinities persisted well into Christian, and even into modern times. Periodic efforts of the medieval Roman Catholic church to halt such worship usually failed and, in fact, led to the association of wells and springs as sites of miracles linked with particular saints (Hope 1893; Hofmann-Krayer and Bachtold-Stäubli 1927; Vaillat 1932; Guitard 1951; Bord and Bord 1986; Guillerme 1988).

Where water cults were restricted to specific springs, it becomes difficult to deal with questions of why springs were worshiped and what the rituals of worship signified. R. A. Wild has argued that the Nile-worshipping cult of Isis and Sarapis (important and widely distributed during the early Roman empire)

simply understood Nile water as the most perfect water; it was associated with fecundity, for humans as well as for crops, and was known as a fattening water (Wild 1981). One may speculate that in a similar sense the worship of the local water source symbolized and represented the dependence of a community on that water. Mineral springs may also have come to be worshiped for their health-giving properties; equally, a spring’s reputation as sacred was an asset to a local economy and made clear to local residents that their locality had a privileged cosmic status (Hamlin 1990a; Harley 1990).

Water as Drink

Water, because it possessed such a broad range of significant and powerful properties, was thus to be used with care in a diet. In the tradition of the Hippocratic writers, authors of medical treatises on regimen had much to say about the conditions of waters and about the circumstances in which water was to be drunk. We should first note that most authors were unenthusiastic about the drinking of water. However much it might seem the natural drink of the animal kingdom, it was also viewed as having a remarkable power to disturb the stability of the human constitution. Summing up the views of classical antiquity on water as beverage, the nineteenth-century chef Alexis Soyer wrote, “water is certainly the most ancient beverage, the most simple, natural, and the most common, which nature has given to mankind. But it is necessary to be really thirsty in order to drink water, and as soon as this craving is satisfied it becomes insipid and nauseous” (Soyer 1977: 299).

The principal later medieval medical text, Avicenna’s *Canon of Medicine*, for example, advised one not to drink water with a meal, but only at the meal’s end, and then in small quantities. Water taken later, during digestion, would interrupt that process. One was also not to drink water while fasting, or after bathing, sex, or exercise. Nor should one give in to night thirst. To do so would disrupt digestion and would not quench the thirst for long. For Avicenna, a water’s temperature was also a crucial determinant of its physiological effect. Too much cold water was harmful, whereas “tepid water evokes nausea.” Warm water acted as a purgative; yet too much of it weakened the stomach (Gruner 1930: 228, 401, 407–8; Lorcin 1985). The effects of habitually imbibing certain waters could be cumulative. The Hippocratic text “On Airs, Waters, and Places,” for example, held that cold waters had a detrimental effect on women’s constitutions: Menstruation was impaired and made painful; breast feeding was inhibited (Hippocrates 1939: 22).

Such a sensitivity to the careful use of water within the diet is also evident in premodern Chinese writings on diet. There, too, one finds recognition of an extraordinary range of properties possessed by different waters (drips from stalactites were seen to

enhance longevity) and, accordingly, great interest in classifying sources of water. In China, the preference was for warm (or boiled) water, possibly, though not necessarily, in which vegetable substances had been steeped (i.e., tea). Cold water was deemed to damage the intestines (Mote 1977: 229–30; Simoons 1991: 24, 441, 463).

Although modes of pathological explanation changed over the centuries, concern with the role of waters in regimens remained important up to the mid-nineteenth century and the onset of a medicine more oriented to specific diseases. The Enlightenment authors of the *Encyclopédie* proposed to determine, through a sort of clinical trial, the full physiological effects of water, but they noted that such a project was impossible because one could not do without water (in some form): One could detect only the differential effects of water and other drinks. (They were particularly interested in the claim that water drinking enhanced male sexual performance [“très vigoureux”]; they thought it probable that such tales reflected only the incapacitative effects of alcoholic drink, not the positive effects of water [Diderot and D’Alembert 1969: entry on “eau commune”]). Late-eighteenth-century British medical men were still stressing the emetic and diluent properties of water. Much food with much water could provoke corpulence; too much water with too little food could promote a diet deficient in nutrients, since food would move too quickly through the digestive tract. For some foods, water was not a sufficient solvent; successful digestion of meats, for example, required fermented beverages (though the alcohol in such beverages was seen as a dangerous side effect). Care was to be taken in quenching thirst, which might not simply signal too little internal moisture but instead indicate too much food or the wrong kinds of food (*Encyclopedia Britannica* 1797: entry on “drink”; *Rees Cyclopaedia* 1819: entry on “diet”).

In antiquity, as in the eighteenth century, the wise physician recognized that general rules like those just mentioned might need modification: One modified them according to a sophisticated explanation of the particular nutritive and other functions of water within the body (further adapted in accord with the constitution and condition of the individual who was to drink it), according to a knowledge of the particular water and the modes of preparation it had undergone. Within the body, water was understood to have a number of effects, both gastronomic and pharmaceutical, but as M. T. Lorcin has noted, in such regimen literature this distinction is inappropriate; health consists of the proper cultivation of the constitution; all ingesta contributed to this (Lorcin 1985: 268).

The chief medical functions of water were as a diluent of food and as a coolant. It acted also as a solvent of food, as an initiator of digestive and other transformations, and as a tonic (a substance that strengthened or gave tone to one’s stomach and/or

other fibers). It was seen also as a mild purgative (Chambers 1741).

The Treatment of Water

To augment some of the functions just mentioned and retard others, waters might be treated or purified. Some of the harmful qualities of water were understood to be susceptible to neutralization or purification. Avicenna recommended boiling as a good means of purification; he held that the mineral residue left behind contained the congealed “coldness” that was the impurity (Gruner 1930: 223). Water might also be made to pass from a container, by capillary action, along a wick of fleece; the drops falling from the end would be assumed to have been purified. Harmful qualities could also be removed by addition of vinegar or wine or by soaking in the water some substance, such as pearl barley, onions, or wax, which would absorb or counteract injurious matters.

One might also shake a suspect water with sand (a technique remarkably similar to that used by Edward Buchner to obtain bacteria-free water for the experiments that led to the concept of the enzyme). Finally, classical and medieval authors recognized the value of filtration, whether a natural filtration through soil, or an artificial filtration through wool, bread crumbs, or cloth (“in order to make sure there are no leeches or other creatures in it”) (Gruner 1930: 222, 454–5; Baker 1948: 1–8; Lorcin 1985: 263). The great potency of water for good and ill, along with its considerable variability from place to place, made it crucial for travelers to be especially careful of the waters they drank: “The traveller is more exposed to illness from the diversity of drinking water than he is from the diversity of foods. . . it is necessary to be particular about correcting the bad qualities of the drinking water, and expend every effort in purifying it” (Gruner 1930: 454).

Thus, long before there was a clear concept of waterborne disease, there was great deal of appreciation, shared by cultures in many parts of the world, of the various characters of waters and of their manifold effects on health. How assiduously people followed hygienic advice about which waters to drink and how to prepare them, and how far following such advice would have been adequate to prevent waterborne diseases is not clear, but it is clear that in cases where public waterworks existed, such as the aqueducts that supplied Rome, those in charge of their administration were supposed to be concerned, in part, with quality. It seems evident that humans have been subject to waterborne diseases throughout recorded history, and, thus, it is remarkable that there is little mention of epidemics (or even cases) of waterborne diseases prior to the nineteenth century (but see Ackerknecht 1965: 24, 41–2, 47, 134–6; Janssens 1983; Jannetta 1987: 148–9; Grmek 1989: 15–6, 346–50).

One might attribute this lack of waterborne epidemics to relatively low population density or a magnitude of travel that was usually too low to sustain outbreaks of diseases caused by relatively fragile bacteria. In this connection, it is notable that many of the records do deal with diseases that we might now attribute to the hardier parasites. Yet it is surely also the case that a population, aware of the dangers of water and possessing an impressive armamentarium of techniques for improving that water, did much to prevent waterborne disease outbreaks. Even if one does not see the Chinese preference for warmed (ideally boiled) water as representing hygienic consciousness, it surely was beneficial in relatively heavily populated areas where paddy cultivation, with night soil as fertilizer, was customarily practiced (but see Needham: 1970). In other cases, as in the addition of wine to water in early modern France, the action was explicitly a purification, with greater or lesser amounts of wine added according to the estimated degree of impurity of the water (Roche 1984).

The Reclassification of Water

Most clearly for Pliny, but also for many classical, medieval, and early modern authors, "waters" were "marvels," each unique, whether owing to the mix of natural agency to which it had been exposed or to its intrinsically marvelous character (Pliny 1963: book 31, chap. 18). By the seventeenth century, European writers on waters had come to emphasize a binary classification: Water was either common (more or less potable) water or mineral water. "Mineral waters" was the collective term for the remarkable springs Pliny had described. Less and less did they represent the mark of the "hand of providence" on a particular locale; increasingly their properties were understood in terms of the salts or gases dissolved in them (Brockliss 1990; Hamlin 1990a, 1990b; Harley 1990; Palmer 1990).

Far from being unique, any mineral spring could be understood as belonging to one of a few general types. These included chalybeate, or iron-bearing waters, drunk to treat anemia; sulfurous waters, good for skin problems; acidulous waters, full of carbonic and other acids that gave the stomach a lightness; and saline waters, which served usually as purgatives. A good many springs, with a wide variety of constituents (and a few with no unusual chemical constituents at all), were also held to be cures for infertility and other diseases of women (Cayleff 1987).

Springs varied in temperature, which might or might not be significant. That waters in some springs were to be bathed in and that waters from others were to be taken internally was a less formidable distinction than it seems to us now. Bathing was not simply a treatment of the skin; the water (or its essential qualities) was understood to be able to enter the body through the skin or to be able to cause signifi-

cant internal effect in some other way. Some therapeutic regimens, such as hydropathy, popular among educated Americans and Europeans in the mid-nineteenth century, integrated a wide variety of external and internal applications of water to produce improvements in health, which clients (like Charles Darwin) regarded as dramatic indeed (Donegan 1986; Cayleff 1987; Vigarello 1988; Brockliss 1990; Browne 1990).

The characterization of springs in terms of chemical constituents was not so much a consequence of the maturation of chemical science as one of the sources of that maturation. Such characterizations were necessary for the proprietors of mineral springs to compete in a medical marketplace. People from the rising middle classes, who increasingly patronized mineral waters, were no more willing to trust in miracles in taking the waters than in any other aspect of business. Every spring had its testimonials, miracles, and claims of excellent accommodations and exalted society for its visitors. Thus, the chemical composition of a spring seemed the only reliable means to make a decision on whether one might patronize a lesser-known resort nearby or whether one undertook a lengthy journey (Guitard 1951; Hamlin 1990b; but see Brockliss 1990). Chemistry also provided a means of bringing the spa to the patient through medicinal waters, which could be bottled for widespread distribution (Kirkby 1902; Coley 1984). Following the discovery of the means to manufacture carbonated water by Joseph Priestly and Torbern Bergman, such enterprise gave rise to the soft-drink industry (Boklund 1956).

One effect, of course, of making water part of the domain of chemistry was to reduce "waters" to a mixture of a simple substrate, "water" (whose composition as a compound of hydrogen and oxygen had been recognized by the end of the eighteenth century) with various amounts of other chemical substances. This conceptual transformation was not achieved without resistance, particularly from physicians (often with practices associated with particular springs), who saw their art threatened by the reductionism of chemistry and continued to maintain that each spring had a peculiar "life" that no chemist could imitate and whose benefits could only be obtained if its waters were drunk on-site (Hamlin 1990a, 1990b).

If, in the light of the new chemistry, mineral waters were no more than mixtures of simple substances, common water was even simpler and less interesting. To medical men and to chemists, this *eau commune* was to be evaluated as belonging to one of two mutually exclusive categories: It was either "pure" or "impure." The terms did not refer to the ideal of chemical purity, which was recognized as practically unattainable; they were simply used to indicate whether the water was suitable or unsuitable for general domestic use (including direct consumption).

At the beginning of the nineteenth century, the

chemists' chief conception of impurity in water was hardness, the presence of dissolved mineral earths. Initially, this new focus supplemented, rather than replaced, the sophisticated classical taxonomy of waters (no one championed soft water that was obviously foul), and the quantification of hardness (expressed as degrees of hardness) was simply a valuable service that chemists could (easily) provide. Hardness was an industrially significant distinction; for steam engine boilers and for brewing, tanning, and many textile processes, the hardness of water was the key criterion. It seemed a medically significant criterion, too: Just like the steam boiler, the drinker of hard water could clog up with bladder stones or gout, conditions that attained remarkable prominence in medical practice, at least in eighteenth-century England (*British Cyclopaedia* 1835: s.v. "water"; Hamlin 1990b).

Waters in the Industrial World

The nineteenth century saw great changes in views of drinkable water and equally great changes in predominant notions of who was competent to judge water quality. Despite the chemists' infatuation with hardness, traditional senses-based approaches to judging water still prevailed in Europe at the beginning of the century. In choosing waters, ordinary people continued to be guided by tradition, taste, and immediate physiological effects. In many cases the standards they used were those found in the classical literature: Stagnant, "foul" water was to be avoided, clear, light, "bright" water was to be desired.

After midcentury, however, expert definitions prevailed. Often experts would insist that water that looked, smelled, and tasted good, and that had perhaps been long used by a local population, was actually bad. Indeed, in some situations, experts' standards were virtually the opposite of lay standards. Light, sharp water had those qualities because it contained dissolved nitrates, which, in turn, were decomposition products from leaking cesspools. The best-tasting well water might, thus, be the most dangerously contaminated (Hardy 1991: 80–1). Less often, experts would insist that waters that laypeople found objectionable (perhaps because they had a strong taste of peat or iron) were wholly harmless.

No longer were chemists restricting themselves to determinations of hardness. Even though the techniques at their disposal did not change significantly (hardness and other forms of mineral content remained the only characteristics they could determine with reasonable effectiveness), chemists increasingly were claiming that they had defined and had the means to quantify what was objectionable in a water beyond its dissolved minerals. They could, they insisted, measure those qualities that had been the basis of the classical water taxonomy better than these qualities could be detected by the senses.

The key quality that interested them, and which at

first supplemented and then displaced concern with hardness or softness, was putridity. "Putridity," while a vague concept, had been the centerpiece of an approach to evaluating water based on one's subjective repugnance to it – owing to its odor, appearance, taste, and associations (the German word *Fäulnis* better embodies such a combination of the visceral and technical). Chemists replaced the senses-based definition of putridity with more arcane indicators of putridity or potential putridity. For much of the nineteenth century, however, they were not in agreement about what precisely these arcane indicators were or the best ways to measure them. Some felt it sufficient to determine the quantity of "organic matter," even though they admitted that this parameter was in some sense an artifact of analytical instrumentation.

Such an approach was contrary to the belief that it was some unknown qualitative factor of organic matter (and not such matter itself) that was associated with putridity and disease. In any case, "foulness" or "putridity" ceased to be a physical state of water and instead became an expert's concept indicating an amount or a presumed condition of "organic" matter. This determination, in turn, was usually believed to correspond to a presumed fecal contamination. Henceforth, the repugnance that "putrid" or "foul" conjured up was to operate through the imagination, rather than directly through the senses.

The champion of this novel perspective was the English chemist Edward Frankland, the leading international authority on water analysis in the 1870s and 1880s. Frankland took the view that it was foolish to try to detect quantities of some unknown disease-generating agency. It was much better simply to try to discover whether water had been subject to contamination in its course through or over the ground. The possibility of dangerous contamination was sufficient reason for public authorities to avoid such supplies of water, and the idea of contamination was to be sufficient to compel ordinary people to avoid its use (Hamlin 1990b).

The shift in approaches to the assessment of water that Frankland exemplifies is a far-reaching one. Associating what might be wrong in water with the presumed commission at some time past of an act of contamination made the religious term "pollution" in its traditional sense of desecration the primary construct for a discussion of water quality, and it came to replace "foul" and "putrid" (Douglas 1966). A presumed act done to the water thus replaced a manifest condition of the water. Although laypersons had known whether water was "foul" at one time, it was up to experts to say whether water had been "polluted." Consequently, water became (and remains) one of very few "foods" whose most important qualities were defined wholly by experts, and whose consumption, accordingly, marked complete trust of the individual in some outside institution: a government, a bottled-water company, or the maker of a filter.

There were, of course, good reasons for such a transformation, and underwriting it was the fear of waterborne (or water-generated) disease. That dense urban environments were dangerous to health was a long-standing medical truth, and water was implicated in this danger: Standing surface water, particularly in marshes, was believed to interact in some way with town filth to generate both fever (particularly malaria) and chronic debilitation. Although consumption of water was not the focus of concern, there was medical consensus that drinking such stuff could not be beneficial to health. Yet at the beginning of the nineteenth century, the doctors were unable to say much about how and in what ways such water was bad, or how it became bad, or how serious a problem bad water might be. Some held that water became harmful by absorbing harmful elements from a filthy urban atmosphere and was simply another means of communicating that state of air. Keeping the water covered would keep it pure, they believed. (Others thought that the putridity was inherent in the water itself and infected the atmosphere.)

Whatever the mechanism, the increasing frequency of epidemic disease was evident in the newly industrialized cities of the nineteenth century. They were swept repeatedly by waves of Asiatic cholera, as well as typhoid fever (only clinically distinguished from other forms of continued fever in the 1840s), and other enteric infections (less clearly identified but no less deadly) (Ackerknecht 1965; Luckin 1984, 1986).

Not until after 1850 were these diseases commonly associated with fecally contaminated water, and even that recognition did not provide unambiguous guidelines for determining water quality because such contaminated water sources only rarely caused severe outbreaks of disease. One might assume that they did so only when contaminated with some specific substance, but as that substance was unknown, it could not be measured, nor was there a clear correlation between the quantity of contamination and the amount of disease. Water that was evidently transmitting cholera was, according to the most sophisticated chemical measures available, substantially purer than water that evidently caused no harm. Some, like Frankland, held that any water that had ever been subject to such contamination should be avoided, but in heavily populated areas, where rivers were essential sources of water, this recommendation seemed impracticable.

Although these contradictions demonstrated the inadequacy of the lay determinations of water quality, the techniques of the experts were little better prior to the twentieth century. Nonetheless, in the nineteenth century, judging waters became a consummately expert task, so much so that the European colony in Shanghai felt it necessary to send water samples all the way to London to be analyzed by Frankland (MacPherson 1987: 85). And even after the microbes responsible for cholera and typhoid were

identified in the 1880s and means were developed for their detection, many experts remained skeptical, unwilling to accept negative findings of their analyses (Hamlin 1990b). But by the early twentieth century, the institution of chlorination, more carefully monitored filtering, and a better understanding of the microbe-removing actions of filters finally led to a widely shared confidence in the safety of urban water supplies (Baker 1948). Such confidence, however, appears to have peaked, and is now in decline.

Town Supplies – Water for All

The fact that cities and towns throughout the world recognize the provision to dwelling houses of piped-in, potable water as an essential component in achieving an acceptable standard of living is remarkable indeed. It involves, in fact, two kinds of public decisions: first, a recognition of a need for a supply of water to be readily available to all settled areas, and second, a recognition of a need for a supply of water to be piped into each dwelling unit.

We usually associate both these features (household water supplies from a public waterworks) as exemplifying the organizational genius of Imperial Rome and lament that it was only in the nineteenth century that authorities, guided by new knowledge of disease transmission and new standards of public decency, again acknowledged water supply as a public duty. Yet a “hydraulic consciousness” was well developed in many medieval and early modern European towns (as well as existing far beyond the Roman empire in the ancient world) (Burton 1868: 241). This consciousness manifested itself in the building of public fountains and pumps, the diversion of brooks for water supply, and even the use of public cisterns and filters, as in Venice. All of these means were used to supply water for industrial purposes, for town cleansing, and for the fighting of fires, as well as for domestic use, but it would appear that the piping of water into individual homes was not felt to be important (Baker 1948: 11–17; Guillerme 1988; Vogel 1988; Dienes and Prutsch 1990; Grewe 1991). High-volume domestic uses of water did not exist; water closets would not become popular until the nineteenth century; bathing was, at least in early modern France, seen to be dangerous to health; and clothing (other than linen) was rarely washed (Vigarello 1988).

By no means was the provision of good drinking water atop this list, but that it was on the list at all is remarkable. How much water people drank, when and where they got it, how public authorities assessed the need for drinkable water and understood their role in supplying it are all questions about which far too little is known. Summarizing pre-nineteenth-century sources, M. N. Baker presented evidence to suggest that urban dwellers did not expect to find raw water drinkable and that knowledge of effective means for treating waters was widespread.

These means ranged from simply allowing sediment in water to settle and then decanting the water to the addition of purifiers (vinegar or wine) or coagulants (alum), or to drinking water only in boiled forms, like tea (Baker 1948: 24–5). The excessive consumption of alcohol was also seen by nineteenth-century temperance advocates as a public response to the unavailability, particularly in poor neighborhoods, of drinkable water (Chadwick 1965: 135–50). It was probably a prudent response: Beer, in particular, was cheap, usually made with a higher-quality water than that readily available, and often more accessible to the poor than water. In Britain it was drunk in hospitals and schools, not just in taverns (Harrison 1971: 37–8, 298–9).

Temperance advocates were often among the champions of public water supplies. But other kinds of reformers became involved, too, sometimes for curious reasons. Those concerned with the morals of the poor worried that a central pump or well was often a locus for the spread of immorality. Children, waiting to fill water containers (sometimes for several hours, if the sources are to be believed), were exposed to bad language, immoral activity, and dangerous ideas. An in-house supply of water could prevent all that. A public drinking fountain movement, begun in Britain in the late 1850s (initially supported by brewers and, surprisingly, not by temperance advocates), received critical support from the Royal Society for the Prevention of Cruelty to Animals, which was concerned about thirsty animals ridden or driven into towns that lacked facilities for watering animals (Davies 1989: 19).

In the century from 1840 to 1940, in almost all of the industrialized world, a public responsibility for providing town dwellers with in-home water was recognized. The timing and circumstances of that recognition varied from society to society (and, significantly, sometimes from town to town), with “public health” considerations usually providing the warrant for that recognition. Adequate sanitary provisions came to include the provision of a water closet in some form and a continuous supply of water that could be drunk without treatment. Whatever its merit on epidemiological grounds, this notion of sanitary adequacy represented the successful promulgation of an ideology of cleanliness and decency that was quite new, and in this transition the status of water changed. No longer was it an aliment whose quality one judged independently for oneself, nor was it something one had to hunt for and sometimes secure only after much labor (Chadwick 1965: 141–2). Instead it was (or was supposed to be) truly a “necessity” of life, something easily and immediately available, nearly as available (and often almost as cheap) as breathable air.

Although the image of water as a public good essential for meeting universal standards of health and decency usually supplied the rationale for under-

taking water-supply projects, the ulterior motives of private interests were often more important in actually getting waterworks built. Perhaps the most significant of these private interests were industrial users. Many industries required large quantities of relatively high-quality water, and the capital costs of obtaining such supplies were prohibitive for individual firms. Consequently, they sought to obtain those supplies (sometimes at subsidized prices) through the sanitary betterment of society. In port cities with much warehouse space, the threat of fire was another underlying incentive for a public water supply.

Some towns took early action to secure control of important watersheds, either with the expectation of profitably selling water to their neighbors or of acquiring commercial advantage. Investors found waterworks projects attractive for a number of reasons, among them a steady dividend, or the possibility of selling land or shares at inflated prices. New York’s first waterworks project attracted speculators because it functioned as a nonchartered and hence unofficial, bank. Some of the capital raised was used to build a waterworks; the rest went into the general capital market (Blake 1956). It need hardly be said that contractors, plumbers, and lawyers were delighted to support waterworks projects. Water is not usually viewed as an article of commerce in the way that most foods are, yet once it had been defined as a public necessity, there was plenty of money to be made from it (Blake 1956; Hassan 1985; Brown 1988; Goubert 1989).

Although this transformation in the availability of water made water drinking much more convenient, the resultant technologies were by no means regarded as an unmitigated benefit. Networks of water mains (and sewer lines) linked people physically across classes and neighborhoods in ways that they had resisted being linked and sometimes in ways that proved hazardous. A common complaint about sewer systems was that they spread disease rather than prevented it because sewer gas frequently rose through poorly trapped drains into houses. It was believed that one was exposed to any infection that the occupants of any other dwelling on the sewer line permitted to go down the drain. More serious, from the perspective of modern epidemiology, was the potential of water mains to distribute infection – precisely what took place in the 1892 Hamburg cholera epidemic (Luckin 1984, 1986; Evans 1987).

Yet for most of the twentieth century, events like the 1892 outbreak of cholera in Hamburg have been rare in the industrialized world. When properly maintained and supervised, the water networks work well. By the end of the nineteenth century, water engineers, finally possessing the torch of bacteriological analysis to illuminate their work, made the filtering of water a dependable operation, even when the water was heavily contaminated. In the first two decades of the twentieth century, they acquired an even more

powerful technique in chlorination. Initially used only when source waters were especially bad or in other unusual circumstances, chlorination quickly became a standard form of water treatment. Even if it was almost always unnecessary, and merely supplementary to other modes of purification, chlorination provided a measure of confidence. That it interfered with (ruined, many might say) the taste of water was no longer of much importance (Baker 1948: 321–56; Hamlin 1990b; O'Toole 1990). Thus, the concept of water as a substance that was necessary to ingest occasionally (even if it was potentially a mode of disease transmission) had, in much of the modern world, very nearly displaced the older concept of "waters" as unique substances, varying from place to place, some of them downright harmful, others with nearly miraculous healthful qualities.

Water in the Present

In recent years, there have been signs that a further transformation of the status and concept of water is under way. In the United States, the authorities responsible for supplying drinkable water are no longer as trusted as they once were (in many parts of the world, of course, such authorities have never known that degree of trust). In some cases that loss of trust reflects a real inability to maintain standards of water quality. But it also reflects public concern about new kinds of contaminants, such as toxic organic chemicals, viruses, and giardia (McCleary 1990; Hurst 1991). In some cases the effects of these contaminants may only be manifest after many years and only through use of the most sophisticated epidemiological techniques (Hand 1988). Nor are customary methods of water analysis or approaches to water purification yet well adapted to such contaminants.

The response of the public has been to revert to a technology, the home water filter (or other water purification devices), that had been popular in the nineteenth century before water authorities were trusted. For some, drinking water has again become a commodity that we think we must go out of our way to secure; something that we haul home in heavy fat bottles from the supermarket. Yet these responses are not adequate to the problem of trustworthiness. The capabilities of domestic water purification devices vary enormously, as does the quality of the product sold by the bottled-water industry and the degree of inspection it receives (Fit to Drink? 1990). Indeed, these responses say less about our need for water we can trust than they do about the institutions we trust.

The rise of the elite bottled mineral waters industry is a reversion too. Pliny tells us that the kings of Persia carried bottled water taken from the River Choapsis with them (Burton 1868: 242); Herodotus and Plutarch referred to an export trade in bottled Nile water – some of it used by devotees of the cult of Isis and Osiris (Wild 1981: 91–4). Such trade was

still widespread in the seventeenth and eighteenth centuries. Then, as now, quality control was a problem and customers complained about the excessive price (Kirkby 1902; Boklund 1956; Coley 1984).

The revival of this industry makes it easier for us to appreciate the fine distinctions among waters made by Pliny, Vitruvius, and the medieval and early modern therapists of the regimen. Modern elites have agreed with their predecessors that the taste (can one say bouquet?) of a water really is important, and that through the drinking of fine waters one can cultivate one's health in ways far more delicate than simply keeping one's insides moist and avoiding cholera.

Christopher Hamlin

Bibliography

- Ackerknecht, E. 1965. *History and geography of the most important diseases*. New York.
- Bachelard, G. 1983. *Water and dreams: An essay on the imagination of matter*, trans. E. Farrell. Dallas, Tex.
- Baker, M. N. 1948. *The quest for pure water: The history of water purification from the earliest records to the twentieth century*. New York.
- Blake, N. 1956. *Water for the cities: A history of the urban water supply problem in the United States*. Syracuse, N.Y.
- Boklund, U. 1956. Tobern Bergman as pioneer in the domain of mineral waters. In *On the Acid of Air*, ed. T. Bergman, 105–28. Stockholm.
- Bord, J., and C. Bord. 1986. *Sacred waters: Holy wells and water lore in Britain and Ireland*. London.
- Brockliss, L. W. B. 1990. The development of the spa in seventeenth-century France. In *The medical history of waters and spas*, ed. R. Porter, 23–47. Medical History, Supplement, 10. London.
- Broughman, J. 1991. The pure truth. *Health* February: 44–6, 57, 102.
- Brown, John C. 1988. Coping with crisis? The diffusion of waterworks in late nineteenth-century German towns. *Economic History* 48: 307–18.
- Browne, J. 1990. Spas and sensibilities: Darwin at Malvern. In *The medical history of waters and spas*, ed. R. Porter, 102–113. Medical History, Supplement, 10. London.
- Burton, R. 1868. *The anatomy of melancholy*. Eighth edition. Philadelphia, Pa.
- Cayleff, S. E. 1987. *Wash and be healed: The water-cure movement and women's health*. Philadelphia, Pa.
- Chadwick, E. 1965. *Report on the sanitary condition of the labouring population of Great Britain*, edited with an introduction by M. W. Flinn. Edinburgh.
- Chambers, E. 1741. *Cyclopaedia or universal dictionary of arts and sciences*. Fourth edition. London.
- Coley, N. 1984. Preparation and uses of artificial mineral waters (ca. 1680–1825). *Ambix* 31: 32–48.
- Davies, P. 1989. *Troughs and drinking fountains: Fountains of life*. London.
- Diderot, D., and D. D'Alembert. 1969. *Encyclopédie ou dictionnaire raisonné des sciences, des arts, et des métiers par une société des gens des lettres*. Elmsford, N.Y.
- Dienes, G. P., and U. Prutsch. 1990. Zur Trinkwasserversorgung von Graz. *Wasser: Ein Versuch*, ed. G. Dienes and F. Leitgeb, 87–90. Graz.

- Donegan, J. 1986. "Hydropathic highway to health": Women and water-cure in antebellum America. Westport, Conn.
- Douglas, M. 1966. *Purity and danger: An analysis of the concepts of pollution and taboo*. London.
- Eliade, M. 1958. *Patterns in comparative religion*, trans. Rosemary Shead. New York.
1978. *The forge and the crucible: The origins and structures of alchemy*. Second edition. Chicago.
- Encyclopedia Britannica*. 1797. Edinburgh.
- Evans, R. J. 1987. *Water in Hamburg: Society and politics in the cholera years, 1830-1910*. New York.
- Fit to drink? 1990. *Consumer Reports*, 27-36.
- Goubert, J.-P. 1989. *The conquest of water: The advent of health in the industrial age*, trans. Andrew Wilson. London.
- Grabner, E. 1990. Wasser in der Volksmedizin. *Wasser: Ein Versuch*, ed. G. Dienes and F. Leitgeb, 237-41. Graz.
- Grew, K. 1991. Wasserversorgung und -entsorgung im Mittelalter - Ein technikgeschichtlicher Überblick. *Die Wasserversorgung im Mittelalter*, 9-86. Geschichte der Wasserversorgung, 4. Mainz am Rhein.
- Grmek, M. 1989. *Diseases in the ancient world*. Baltimore.
- Gruner, O. C. 1930. *A treatise on the canon of medicine of Avicenna incorporating a translation of the first book*. London.
- Guillermé, A. E. 1988. *The age of water: The urban environment in the north of France, A.D. 300-1800*. College Station, Tex.
- Guitard, E. H. 1951. *Le prestigieux passé des eaux minérales*. Paris.
- Hamlin, C. 1990a. Chemistry, medicine, and the legitimization of English spas, 1740-1840. In *The medical history of waters and spas*, ed. R. Porter, 67-81. Medical History, Supplement, 10. London.
- 1990b. *A science of impurity: Water analysis in nineteenth century Britain*. Berkeley, Calif.
- Hand, D. 1988. Poisoned water: The long struggle to be heard. *America's Health* May: 48-53.
- Hardy, A. 1991. Parish pump to private pipes: London's water supply in the nineteenth century. In *Living and dying in London*, ed. W. F. Bynum and R. Porter, 76-93. Medical History, Supplement, 11. London.
- Harley, D. 1990. A sword in a madman's hand: Professional opposition to popular consumption in the waters literature of southern England and the midlands, 1570-1870. In *The medical history of waters and spas*, ed. R. Porter, 48-55. Medical History, Supplement, 10. London.
- Harrison, B. 1971. *Drink and the Victorians: The temperance question in England, 1815-1872*. Pittsburgh, Pa.
- Hassan, J. A. 1985. The growth and impact of the British water industry in the nineteenth century. *Economic History Review* 38: 531-47.
- Hesiod. 1953. *Hesiod's theogony*, trans. Norman O. Brown. New York.
- Hippocrates. 1939. On airs, waters, and places. In *The genuine works of Hippocrates*, trans. F. Adams. Baltimore, Md.
- Hofmann-Krayer, E., and H. Bachtold-Stäubli, eds. 1927. Brunnen. In *Handwörterbuch des deutschen Aberglaubens*. Berlin.
- Hope, R. C. 1893. *The legendary lore of the holy wells of England: Including rivers, lakes, fountains, and springs*. London.
- Hurst, C. J. 1991. Presence of enteric viruses in freshwater and their removal by the conventional drinking water treatment process. *Bulletin of the World Health Organization* 69 (1): 113-19.
- Jannetta, A. 1987. *Epidemics and mortality in early modern Japan*. Princeton, N.J.
- Janssens, P. A. 1983. The morbus desentericus in the *historia francorum* of Gregory of Tours (sixth century). *Disease in ancient man: An international symposium*, ed. G. Hart, 263-6. Toronto.
- Kirkby, W. 1902. *The evolution of artificial mineral waters*. Manchester, England.
- Lorcin, M. T. 1985. Humeurs, bains, et tisaines: L'eau dan médecine médiévale. *L'Eau au Moyen Age*, 259-73. Publications du Cier Ma, Université de Provence. Marseille.
- Luckin, B. 1984. Evaluating the sanitary revolution: Typhus and typhoid in London, 1851-1900. *Urban disease and mortality in nineteenth-century England*, ed. R. Woods and J. Woodward, 102-19. New York.
1986. *Pollution and control: A social history of the Thames in the nineteenth century*. Bristol, England.
- MacPherson, K. 1987. *A wilderness of marshes: The origins of public health in Shanghai, 1843-1893*. Hong Kong.
- McCleary, K. 1990. Trouble from your tap. *Health* May: 32-3, 76.
- Moser, H. 1990. Brunnen, Bründl und Quellen: Eine volkskundliche Betrachtung. In *Wasser: Ein Versuch*, ed. G. Dienes and F. Leitgeb, 234-6. Graz.
- Mote, F. 1977. Yuan and Ming. *Food in Chinese culture: Anthropological and historical perspectives*, ed. K. C. Chang. New Haven, Conn.
- Needham, J. and L. 1970. Hygiene and preventive medicine in ancient China. In *Clerks and craftsmen in China and the west*, ed. J. Needham, 340-78. Cambridge.
- O'Toole, C. K. 1990. *The search for purity: A retrospective policy analysis of the decision to chlorinate Cincinnati's public water supply, 1890-1920*. New York.
- Palmer, R. 1990. "In this our lightye and learned tyme": Italian baths in the era of the Renaissance. In *The medical history of waters and spas*, ed. R. Porter, 14-22. Medical History, Supplement, 10. London.
- Partington, Charles, ed. 1835. *British cyclopaedia of the arts and sciences*, 885-95. London.
- Paulus Aegineta. 1844. *The seven books of Paulus Aegineta with a commentary*, trans. F. Adams. London.
- Pliny. 1963. *Natural history with an English translation in ten volumes*, trans. W. H. S. Jones. Cambridge, Mass.
- Rees, Abraham. 1820. *The cyclopaedia; or Universal dictionary of arts, sciences, and literature*. London.
- Roche, D. 1984. *Le temps de l'eau rare: Du moyen age à l'époque moderne*. Annales E.S.C. 39: 383-99.
- Simoons, F. J. 1991. *Food in China: A cultural and historical inquiry*. Boca Raton, Fla.
- Soyer, A. 1977. *The pantropeon or a history of food and its preparation in ancient times*. New York.
- Taylor, J. G. 1982. *Eating, drinking, and visiting in the south: An informal history*. Baton Rouge, La.
- Tölle-Kastenbein, R. 1990. *Antike Wasserkultur*. Munich.
- Tuan, Y. F. 1968. *The hydrologic cycle and the wisdom of god: A theme in geoteleology*. Toronto.
- Vaillat, C. 1932. *Le culte des sources dans la gaule antique*. Paris.
- Vigarello, G. 1988. *Concepts of cleanliness: Changing attitudes in France since the Middle Ages*, trans. J. Birrell. Cambridge, New York, and Paris.
- Vitruvius Pollio, Marcus. 1960. *The ten books on architecture*, trans. M. H. Morgan. New York.
- Vogel, H. 1988. Brunnen und Pumpereien in der Stadt Bremen. *Wasser. Zur Geschichte der Trinkwasserver-*

sorgung in Bremen. Ausstellungskatalog Bremer Landmuseum für Kunst - und Kulturgeschichte, 50-67. Bremen.

Wild, R. A. 1981. *Water in the cultic worship of Isis and Sarapis*. Leiden, Netherlands.

III.13 🍷 Wine

Wine is the fermented juice (must) of grapes, and for thousands of years humans have been attempting to perfect a process that occurs naturally. As summer turns into fall, grapes swell in size. Many will eventually burst, allowing the sugars in the juice to come into contact with the yeasts growing on the skins. This interaction produces carbon dioxide, which is dissipated, and a liquid containing alcohol (ethanol) in combination with a plethora of organic compounds related to aroma and taste that have yet to be fully enumerated. Many people have found drinking this liquid so highly desirable that they have been willing to expend enormous effort to find ways of improving its quantity and quality.

In some places both viticulture (grape growing) and viniculture (wine making) emerged as specialized crafts, which today have achieved the status of sciences within the field known as enology. In general, three basic types of wine are produced: (1) still or table wines with alcohol contents in the 7 to 13 percent range; (2) sparkling wines from a secondary fermentation where the carbon dioxide is deliberately trapped in the liquid; and (3) fortified wines whereby spirits are added to still wines in order to boost their alcohol contents into the 20 percent range.

Vine Geography

Grape-bearing vines for making wine belong to the genus *Vitis*, a member of the family Ampelidaceae. Vines ancestral to *Vitis* have been found in Tertiary sediments dating back some 60 million years, and by the beginning of the Pleistocene, evolution had produced two subgenera - *Euvitis* and *Muscadinia*. Both were distributed across the midlatitude portions of North America and Eurasia. Glaciation, however, exterminated the Muscadines with the exception of an area extending around the Gulf of Mexico and into the southeastern United States, where one species, *Vitis rotundifolia*, has been used to make sweet wines that go by the regional name of scuppernong.

The distribution of wild *Euvitis* in North America was concentrated east of the Rocky Mountains. In excess of 30 species have been identified, with *Vitis labrusca* being by far the most important in terms of wine making. Its genes are found in all the major grape cultivars such as the Concord, Niagara, Catawba, Dutchess, Delaware, and Elvira. Other species that

have made some significant contributions include *Vitis riparia*, *Vitis aestivalis*, and *Vitis bourquiniana*.

Across Eurasia, the geography of *Euvitis* reflects the cultural importance of wine. From southwestern Asia through Europe one species, *Vitis vinifera*, dominates, but from central Asia to Japan and Southeast Asia, where drinking wine made from grapes is not a traditional practice, a number of species still can be found. The wine-making superiority of *V. vinifera* undoubtedly led to its replacing competing species. It achieves a balance between sugars and acids that is unmatched by any of the others, and it evolved with human assistance into a self-propagating hermaphrodite capable of producing a virtually unlimited number of cultivars possessed of different colors, aromas, and taste characteristics. Some of the better known of these include Cabernet Sauvignon, Pinot Noir, Chardonnay, Riesling, Syrah, and Nebbiolo, but there are many hundreds of others, most with highly localized distributions.

In the late nineteenth century, French scientists began systematically to hybridize *vinifera* with *labrusca* and other native American vines. Their efforts were aimed at producing varieties that were more disease resistant (ones that could, in particular, resist the two scourges of phylloxera and oidium) and were better able to withstand winter cold. Often called French-American hybrids, they were planted in many parts of Europe during the early twentieth century, but currently most of the vines are being uprooted because of dissatisfaction over their wine-making capabilities. In the eastern United States and Ontario, Canada, however, the hybrids have become increasingly important, and research continues into developing new varieties.

During the last 500 years, humans have carried *Euvitis* to areas outside its natural range. It is now found growing in the tropics, as well as in the midlatitude portions of Chile, Argentina, South Africa, Australia, and New Zealand, where wine making has achieved considerable prominence. Virtually all of the vineyards are planted with *V. vinifera*.

Early History

The earliest archaeological evidence indicating wine that might have been made from domesticated vines comes from a pottery jar, dated to between 7400 and 7000 years ago, which was found at the Neolithic site of Hajji Firuz in the northern Zagros Mountains. Whether this is wine's precise hearth remains for future research to determine. But the general area from the Zagros to the Caucasus Mountains is a reasonable possibility - it lies within the natural range of *V. vinifera* and it is included within the broader Southwest Asian region of plant and animal domestication. From here, wine seems to have diffused in two directions.

One was into Assyria and thence to the Mesopotamian city states of Kish, Ur, and later Babylon, among which it attained the status of a luxury

drink meant for priestly and noble classes. The hot, dry, and saline soil conditions of Mesopotamia were not well suited for growing grapes, and thus most of its wines probably were imported from the north.

As with many other facets of culture, wine seems to have entered Egypt from Mesopotamian sources. Vines were established along the Nile by at least 5000 years ago, and during New Kingdom times a highly sophisticated technology of wine growing was in place, including an appreciation of vintages and aging. Wine growing followed the Nile upstream to Nubia, where it persisted until Islam's hegemony over the area was established early in the second millennium A.D.

By 3000 years ago, wine drinking had become widespread in southwestern Asia. It was found among the Canaanites and the closely related Phoenicians, and wine consumption became customary among the Hebrews, who seem to have employed it as a ceremonial drink that was taken in moderation by everyone in the community. Elsewhere, the evidence indicates that wine continued to be restricted to those at the top of the socioeconomic ladder and that it was appreciated largely for the intoxicating effects of its alcohol. Intoxication, indeed, was undoubtedly the reason why almost from its inception wine and religion were bound closely together – wine served as a medium for contacting the gods.

The second route carried wine across Anatolia to the shores and islands of the Aegean. It was adopted by the Minoan civilization on Crete and in Mycenaea, where its production became a specialization of some islands and an important item of commerce. Most significant for later developments, wine became a part of Greek culture, and wherever the Greeks planted their colonies in the Mediterranean, wine drinking, wine growing, and wine commerce followed. Sicily and southern Italy became especially prominent and were given the name *Oenotria*, the land of vines. Another important site was Massilia (Marseilles), from where the Greek method of wine growing spread inland following the Rhone River valley.

Wine became even more popular because of the cult of the Greek god Dionysus (Bacchus to the Romans). Originating as a minor fertility symbol among the Anatolians, he was transformed into a complex figure representing wine in the context of religious and political protest. These protests often took the form of ecstatic revelries, aided by drunkenness, that were intentionally designed to violate prevailing behavioral norms. The cults spread widely throughout the Mediterranean world and apparently were highly instrumental in creating a greater demand for wine among the populace at large.

The Roman Factor

The Romans are usually credited with having contributed more to the advance of wine than any other people. In fact, they probably were not wine drinkers,

or at least not wine growers, when they first embarked on their expansion, but instead acquired the habit from the Etruscans, a people derived either from Minoan or Mycenaean migrants who settled in central Italy. But there is little doubt that wine growing flourished under Roman influence. It was established virtually everywhere on the Italian peninsula, both on large noble-owned estates and as a part of peasant farming systems.

Wine also expanded with the empire. A nearby source of supply was needed for garrisons and administrators, and within Rome itself there was a continuous demand for new wines among the upper classes. Although archaeological evidence indicates that some of the Celtic, Iber, Germanic, and Slavic peoples knew how to make wine, they appear to have been largely beer or mead drinkers. Wherever the Romans imposed their rule, however, fundamental changes in wine's regional significance followed. The Romans not only brought with them a wine-drinking tradition that was copied by others but also set about systematically identifying and developing those areas best suited to vineyards. Key considerations included a nearby market, usually an important town, a riverine or coastal location to facilitate transport, and where possible, slopes for correct exposure to the sun and for facilitating the flow of air and soil moisture drainage.

The greatest amount of Roman wine-growing attention was focused on Gaul and the Rhineland. Today, virtually every important wine region in France and Germany, such as Champagne, Burgundy, Bordeaux, and the Rhineland (and many lesser ones as well) can find prominent Roman influences in its history. Viticulture was a particular specialty of the Romans; they created numerous new grape varieties that were better adapted to more northerly sites and they improved methods of vine cultivation, especially techniques related to pruning and training. Although there is, perhaps, a tendency to credit the Romans with too much in many facets of culture and technology, it is quite clear that the map of wine growing would have been drawn much differently without them.

Decline and Rebirth

The demise of Roman political and military authority led to a reversal of wine's fortunes in Europe. This was not because successors, such as the Franks, Vandals, Goths, Visigoths, and Huns, eschewed wine; rather it resulted from the unsettled conditions that plagued the countryside for several centuries. Vineyards were difficult to maintain in the face of repeated raiding and pillaging, and markets all but disappeared as cities went into decline with the disruption of trade. Wine growing persevered, if on a much reduced scale, and it managed to do so largely because of its role in the Christian liturgy. Wine was needed in order to celebrate the mass, and thus the

numerous bishoprics established following Rome's conversion managed to keep small vineyards in production during the difficult times of the middle of the first millennium A.D.

Of longer-term significance, however, were the monasteries, which provided bastions of security amid general insecurity. In terms of wine, the most significant monastic orders were the Benedictines and their offshoot the Cistercians, followed by the Carthusians. Experiments by the monks led to new cultivars and to improvements in numerous aspects of viticulture, especially with regard to site selection. Many of today's most celebrated vineyards in such places as Burgundy, the Reingau, and the Mosel were laid out by monks nearly a thousand years ago.

By contrast, from Persia through western Asia across northern Africa and into Iberia, viticulture was confronted by the advance of Islam and the Koranic injunction against drinking alcohol in any form. Although the spread of Islam was very swift, the elimination of wine growing proceeded much more slowly. Minority Christian and Jewish communities often were allowed to continue to use the beverage largely for the tax revenues wine provided. Thus during the early centuries of Islamic rule, the ban on alcohol varied in terms of how rigorously it was enforced. It never seems to have taken a strong hold in Iberia, and consequently wine growing continued there without serious interruption during the period of Moorish rule. But wine in western Asia and the eastern Mediterranean was under greater pressure, and with the rise to power of the Ottomans, it virtually disappeared over much of the area.

Although Norse raids had slowed the pace of vineyard developments in Europe, expansion was once more under way by the beginning of the twelfth century, even in such northerly locales as the Low Countries and England. The monasteries continued with their plantings and were joined in this endeavor by noble estates, many of which employed wine-growing specialists. Local demand was expanding as wine, at least in Western Christendom, became a more common drink among the populace at large. It was a substitute for unsafe drinking water and widely prescribed as a medication for a variety of ailments. In addition, profits could be gained from exports, particularly to the trading states of the Baltic and North Seas. Much of the wine headed in this direction seems to have come from the Rhineland, including Franconia and Alsace.

Markets and Politics

By the close of the twelfth century, the history of wine growing in Europe was being shaped to an ever increasing extent by a combination of politics and market demands in which England played the most prominent role. Vineyards had been planted in the southern part of that country during Roman times,

and as noted, further developments occurred with the establishment of monasteries.

The Domesday Book recorded in excess of 40 vineyards, and over the course of the next several centuries their numbers multiplied severalfold. Most, however, seem to have been quite small, and the wine they produced never achieved any esteem. It was thin and acidic and used primarily for sacramental purposes. Thus the English had to begin looking elsewhere for the wine they were consuming in ever larger volumes. The future of wine growing in England was further dimmed by the advancing cold of the Little Ice Age. By the seventeenth century, all the English vineyards had fallen into disuse.

Much of the wine the English initially imported came from the Rhineland, but the source of supply shifted to the Loire River Valley and its hinterlands when Aquitaine was joined to England during the middle of the twelfth century. An important commercial center in Aquitaine was Bordeaux, and here a new kind of wine was encountered. Unlike the white to golden-colored wines from the Rhineland and the Loire, the wine here was red, or at least pink. Termed claret, it became the royal favorite and thus the wine in most demand among the upper classes. Yet at this time, only a very small amount of claret actually came from the Bordeaux region itself. Most of it was shipped down the streams converging on the Gironde estuary from the *hauts pays* of the interior, from such places as Bergerac, Gaillac, and Cahors.

In the mid-fifteenth century, following the end of the Hundred Years' War, all of Aquitaine, including Bordeaux, was incorporated within France. This development did not end the flow of claret to England, but the supply became less regular and depended to a great extent on the state of relations between the two countries. English merchants looked for alternatives, which took them southward to Portugal, their political ally against the Spanish and French.

Some red wines were exported from the Minho region of Portugal, but they never developed much of a following. As wars and plans for war grew more frequent toward the close of the seventeenth century, the search for reliable supplies of claret substitutes intensified. Some merchants traveled up the rugged and underdeveloped Duoro Valley from their base at Porto. Only very rough reds were encountered, and their initial reception in England was not very positive.

What caught on eventually, however, was the practice of adding brandy to the still wines. At first this was done at the end of fermentation to increase alcohol content. High-alcohol wines were preferred not only because they were more quickly intoxicating, but also because they did not spoil as rapidly and thin into vinegar. Although not discovered until much later, the wine-spoiling culprits were oxidation and bacterial contamination, and the reactions they produced slowed down at higher alcohol levels.

In the eighteenth century, it was discovered that if brandy was added during fermentation, a sweet wine of much higher quality could be produced, and thus the port of today was born, named for Porto, the place where it was processed and from which it was shipped. It became a highly fashionable drink for English gentlemen, which set off a wave of vineyard plantings that transformed the upper Duoro Valley into a landscape almost totally devoted to wine.

Perhaps the most highly prized wine in England was malmsey. It arrived in Genoese and then Venetian galleys from Cyprus and other Christian-held lands in the eastern Mediterranean, where vineyards had been reestablished by monastic orders in the wake of the Crusades. These sun-drenched lands produced a sweet, golden wine naturally high in alcohol, and thus the malmsey also traveled well. However, the Ottoman advance cut off the source of supply, which once again set in motion a search for alternatives that eventually led to Spain. A wave of vineyard plantings there had accompanied the Reconquest, and the expulsion of the Jews had provided an opportunity for English merchants to set themselves up in business, particularly at the active ports of Cadiz and Sanlúcar. The wine exported from this southern coastal region was initially called sac, and later sherry. The latter was an English corruption of the word Jerez, the town in which most of the wine was made. Like the port wines, sherries were fortified and initially sweet. Later developments led to dry styles gaining preference, with England maintaining an almost exclusive hold on the market. A similar type of wine was produced on the island of Madeira, but because of the papal division of the world, its exports were directed mostly to the Americas.

The other major participant in the wine trade was Holland, where cool growing-season temperatures, poorly drained soils, and high winds virtually precluded the establishment of vineyards. Satisfying most of the Dutch thirst for wine were the vineyards of the Rhineland and the Mosel Valley, but the destruction wrought there by the Thirty Years' War caused the Dutch (like the English) to turn to the Loire Valley and to Bordeaux, where they encouraged the growing of grapes to make white wines.

Dutch tastes, however, were increasingly moving in the direction of spirits, especially brandies. In addition to vineyards, ample timber supplies were required in order to fuel the distilleries, and the Dutch found this combination in two locales, the Charente (Cognac) and Armagnac, two names that have come to define the standards by which brandy is judged.

The Search for Quality

The emergence of noble classes and wealthy merchants in Europe generated a greater concern for wine quality, with those in Burgundy leading the way. There, in the fourteenth century, the combination of

the tastes of the dukes and the wine-growing skills of the Cistercians resulted in a purposeful attempt to reduce the quantity of wine produced in order to raise its quality. The monks had determined that the best vineyard sites lay along a faulted escarpment that has since gained the name Côte d'Or, and a Duke, Philippe the Bold, declared the high-yielding Gamay grape banned in favor of the lower-yielding but more aromatic and complexly flavored Pinot Noir (then known as the *noirien*). At about the same time, quality considerations were singling out the Riesling grape in the Rhineland and the Mosel Valley. Both cases illustrate an important trend in the making – the development of a hierarchy of quality wine regions based in large measure on grape varieties.

Two advances of the seventeenth century were crucial in the quest for wine quality. The first was the discovery of how to make stronger glass bottles, thus providing an inert container in which wine could be stored instead of in wooden barrels, jugs, or, in some instances, leather bags. The bottle then needed a stopper to protect the wine from the deleterious effects of too much oxygen contact. Wood plugs and cloth were tried with limited success, but the solution to the problem proved to be cork. It deterred the passage of air, and cork's pliability meant that it could be made to fill the space between the walls of the bottle's neck. It was now possible to keep new wines fresh longer, and a medium was available to allow wines to age.

One region that profited almost immediately from the use of improved bottles and corks was Champagne. By the fourteenth century, its wines had gained the reputation of being good values. Champagne was close to the Paris market, and reds and whites were made in styles similar to those in Burgundy. Matters began to change, however, in the seventeenth century under the leadership of the Benedictine abbey at Hautvilliers and, most particularly, when the legendary Dom Pérignon assumed the position of cellar master. He set meticulous cultivation and vinification standards designed to elevate Champagne's wines to the very top of the quality ranking. Consistency seems to have been what he was seeking most, and to this end he perfected the art of wine blending, something that has been a hallmark of champagne ever since.

One of the problems Dom Pérignon had to deal with was a fizziness that frequently appeared in the wine during the spring following vinification. The cold weather of the late fall sometimes shut down fermentation before it was completed, and it would start up again when the warmer weather of spring arrived. Yet the fizziness produced by the carbon dioxide in the wine was precisely what consumers began to find attractive. It became fashionable to serve this kind of champagne at the court of Louis XIV during festive occasions, and the habit soon spread to other European royal families, setting off a demand that could

not be met by relying on an accident of nature. Thus, by the early part of the eighteenth century, secondary fermentation was being purposefully created through the addition of a dose of yeast and sugar to already bottled still wine. Because the carbon dioxide created enormous pressure, strong bottles and a tight seal were both required. During the remainder of the eighteenth century and on through the nineteenth, numerous technological changes continued to improve champagne's quality and thereby its image. It was a different wine than that which Dom Pérignon envisioned, but it achieved the ranking he sought.

No region is more closely associated with the image of wine quality than Bordeaux. As noted, the politics of the seventeenth and eighteenth centuries meant that the supply of claret for the English market was highly unreliable, and often it fell far short of demand. Consequently, claret became a prestigious commodity with attendant high prices. This situation initiated a twofold response from Bordeaux. One was an effort by some producers to make better wines and then to let consumers know with a label which properties or estates had produced them. The wine was thus not just a claret but a claret from this or that estate. The other was to expand markedly the amount of vineyard land. Most of the plantings initially were near the city of Bordeaux, in an area known as the Graves because of its gravelly soils. Because of drainage work carried out by Dutch engineers earlier in the seventeenth century, a considerable amount of land was available in the Médoc promontory to the north of the city, and by the end of the century, most of the best sites, also on gravel, had been planted. Expansion continued during the next century and soon enough wine was being produced in Bordeaux so that supplies were no longer needed from the *hauts pays*. Indeed, these wines were restricted from entering in order to protect local prices.

Wine and wealth in Bordeaux had come to feed on one another. The port was significant in its own right, connecting the southwest of France not only to England but also to the markets of Scotland, Ireland, Holland, and the northern German states. As profits grew, more and more was invested in wine, both by merchants and the landed aristocracy. The latter began to build country estates marked by grand chateaux, which helped to further the elite image of the Bordeaux wines. In terms of prices paid and profits earned, Bordeaux had risen to the top of the wine world.

Imperial Expansion

Wherever possible, wine growing accompanied European imperial expansion. It was introduced into Mexico during the first decades of the sixteenth century and from there spread southward in rapid order to Peru, Chile, and Argentina. Its northward expansion was slower, but eventually by the late eighteenth cen-

tury it had reached what is now California. Catholic missions had established the first vineyards in both directions. They were small, and the wine produced went mainly to celebrate the mass, but some was also consumed at meals by priests and monks and by nearby Spanish settlers. Wine, however, was not to be given to the local Indians.

Where and when wine was first made in the eastern part of the United States remains in dispute. Some give the credit to French Huguenots at a settlement near Jacksonville, Florida, in the 1560s; others point to the settlers of a contemporaneous Spanish colony on Santa Elena Island off the coast of Georgia; and still others say it was the Jamestown settlers. Whoever they were, they would have used wild native American grapes, including the muscadines. The pungency of wines made from these grapes was not greatly appreciated, and throughout the colonial period attempts were made to grow *vinifera* varieties. Without exception, however, the vineyards died within a few years of planting (we shall see why in the next section). Reconciling themselves to this fact, wine makers turned their attention toward domesticating and improving the indigenous vines, the first of which, known as the Alexander, appeared at the end of the eighteenth century.

Vinifera found a much more hospitable environment at the very southern tip of Africa. Here the Dutch East India Company chose Table Bay as the site for establishing a colony in 1652 to revictual their ships rounding the Cape of Good Hope, and vines were among one of the first crops to be planted. In the late 1680s, a group of French Huguenot refugees arrived; they brought wine-growing expertise with them and extended the area devoted to vineyards. The vines did so well in the mild Mediterranean climatic conditions of the Cape region that not only were the ships resupplied with ample quantities for their crews, but exports were also sent to Europe. One result was a rich dessert-type wine known as Constantia that gained considerable renown in the early 1700s.

Wine growing was attempted by the first Australian settlers who arrived at Sydney Harbor in the 1780s using seeds and cuttings obtained from Europe and the Cape. But the prevailing high heat and humidity militated against success and, in any event, more money could be made raising sheep. It was not until the 1820s that wine growing in Australia finally found a congenial environment in the Hunter Valley. Other areas in Victoria, southern Australia, and western Australia came into production in the 1830s and 1840s when wine-making efforts were also occurring on the North Island of New Zealand.

Disasters in the Making

By the middle of the nineteenth century, wine was entering what many have termed its golden age. It had been an everyday drink among Mediterranean

Europeans for many centuries and was becoming ever more common across the Alps, where a succession of good vintages helped to spur an increase in the areas given over to vineyards. Profits had never been greater, and some areas, notably the Midi of France, became virtual monocultures supplying the demands of the general public. Side by side were other areas catering to the wealthy, now augmented by newly emergent industrial and professional classes. Burgundy, Champagne, and the Rhine/Mosel each had its markets, but no place could challenge the primacy of Bordeaux.

Most of the wine being produced was, in fact, of ordinary quality, but a few properties, mostly located in the Médoc, had managed to distinguish themselves. This led to numerous attempts at classification, the official codification of which was achieved in 1855 for the Paris Universal Exposition. At the top, it divided the wines of the Médoc into first through fifth growths (*crus*), and it then added lower rungs of *crus exceptionnelles*, *crus bourgeois*, and *crus artisans*. The appearance of such a listing, which no other wine region had, created an even greater demand for the top growths, and Bordeaux's image reached new heights.

Nevertheless, disaster was lurking in the background. It started with a powdery mildew, termed oidium, that in the early 1850s was infesting vineyards virtually everywhere in Europe. The fungus retards vine growth by attacking both leaf and stalk, and it also desiccates grapes by breaking their skins. Consequently when it first appeared, yields declined, as was especially the case from 1852 to 1854. Nevertheless, a treatment in the form of sulfur dusts was quickly found, and by the end of the decade oidium, although not eliminated, was being controlled.

But no sooner had oidium come under control than reports of mysterious vine deaths began to emerge. The first came from the lower Rhone Valley, and by 1870, it had become evident that the problem was on the verge of universality. The plague of vine destruction made its way through France (the last region to be visited was Champagne in 1901) and the rest of Europe, and eventually into most of the wine-growing world. Chile was the only major exception, being spared by a combination of distance, desert, and mountains.

In 1868 the culprit was shown to be a minute aphid, *Phylloxera vastatrix*, that destroys vine roots. It is native to North America, and may have entered France on vines that were imported for experiments on ways to better control oidium. With the collapse of the wine industry at hand, numerous, often desperate, attempts were made to rid vineyards of the aphids. Some chemical treatments were moderately effective, but the ultimate solution turned out to be one of grafting European vines onto American roots. Experiments had determined that several species of the latter were tolerant of the aphid's presence and thus

North America, which was the cradle of phylloxera, turned out to be the cradle of the cure.

But the diffusion of phylloxera had drastically altered the wine-growing world. Those regions hit later were able to realize short-term profits. For example, the price of champagne soared, and Italy and Spain filled in the gaps left by the fall of French production in the 1880s and 1890s. In addition, wine growers themselves were on the move. Algeria was a favorite destination for those leaving France, and small numbers also relocated in the Rioja region of Spain, bringing with them new ways of growing vines and making wine with an eye on the larger market. When, at the turn of the century, phylloxera finally hit hard in Italy, Spain, and Portugal, an exodus to Chile, Argentina, and Brazil ensued, leading to the reenergization of these long dormant regions.

Yet a third natural disaster had to be faced, that of downy mildew, which also found its way to Europe on American vines. In a manner similar to oidium, it reduced crop yields and required dusting, this time with a mixture of copper sulfate and lime, to keep it under control.

The labor and cash requirements of grafting and repeated sprayings drove many farmers in Europe out of the wine business, which was no longer something that one could do in a small way. Rather it required new skills and greater financing in order to be competitive, and these were best obtained by specialization. Virtually everywhere, the making of wine was rapidly changing from a craft into a business and a science.

In the United States, oidium, phylloxera, and downy mildew had also taken their toll. Indeed, it was phylloxera that had been responsible for the repeated failure of *vinifera* to survive east of the Rocky Mountains, and in the 1850s an outbreak of oidium had put an end to the country's first wine region of note – the so-called Rhine of America along the Ohio River near Cincinnati. A far more significant disaster, however, was social and political in nature, namely Prohibition. Its origins go back to the temperance crusades of the nineteenth century designed to purify America of all vices, and most especially those thought to be associated with alcoholic beverages. The purification urge culminated in 1919 in passage of the Eighteenth Amendment to the U.S. Constitution (popularly known as the Volstead, or Prohibition, Act), and its 13 years of existence brought nearly total ruin to the country's wine industry. By this time, California had achieved top rank in terms of both quantity and quality, but wineries abounded in many other states. Very few managed to survive Prohibition. Those that did sold sacramental and medicinal wines, as well as grape concentrate for making wine at home, all of which could be done under provisions of the Volstead Act.

Over most of the first half of the twentieth century, the world's wine industry remained in a

depressed state. In addition to the ravages wrought by disease and Prohibition in America, the industry had to contend with two world wars, the Russian Revolution (which eliminated an important market for French wines, particularly champagne), and economic depression. In retrospect, the golden age truly glittered.

Recovery and Transformation

Recovery was fueled by the widening base of prosperity of the immediate post-World War II years and by a Europeanization of food habits in many parts of the world. Outside of Europe, wine drinking tended to be relatively uncommon even in those areas settled by Europeans. In general, members of the working and middle classes drank mainly beer and spirits or abstained from alcohol altogether. However, by the 1950s more people could afford wine, and many discovered it as an enjoyable accompaniment with dinner. For others, wine served as a symbol of their upward mobility.

The French were the first to prosper. They had solidified their image as the world's quality leader by developing a legal classification of wines known as *Appellation d'Origine Contrôlée* (AOC). Its outlines were formed early in the twentieth century in order to counter an outbreak of fraud following the wine shortages caused by phylloxera. Fraudulent practices included such things as the marketing of wine from one region under the name of another; placing non-French wines under French labels; and fermenting a range of fruits and other substances in place of grapes.

As the AOC system evolved into its present form, the emphasis was placed on guaranteeing the authenticity, especially the geographic authenticity, of wines. The system defines specific regions, dictates where the grapes that go into the wine of those regions must come from, and also requires adherence to precise viticultural practices that are deemed essential to achieving a wine's particular character. A certain level of quality is implied, but only in a very few instances are taste tests actually performed. The success of the AOC system in heightening the image of French wines meant that other countries began creating their own versions. Each is constructed differently, but the message is basically the same – the government certifies what is in the bottle.

Wineries in California led to other trends. Of special significance was the state's abilities to produce large volumes of consistently good-quality wines at relatively inexpensive prices. Three related factors made this achievement possible. One was the high grape yields attainable in the rich, irrigated central valley of the state; another involved the development of new quality-control technologies, especially ones designed to control fermentation temperatures, permitting bulk manufacturing; and the third was the for-

mation of large companies that could afford to adopt these technologies as well as to engage in mass advertising campaigns. Thus, while meeting a demand, the companies were also creating it.

At the same time, other California producers were heading in a somewhat different direction. Following the repeal of the Volstead Act, several smaller-volume wineries in the Napa Valley north of San Francisco had decided to compete with French wines in the prestige market. Their success in the 1950s and early 1960s sparked a wave of winery openings that spread out from the Napa Valley to adjacent Sonoma County and then to other areas of the state. Leading both branches of the industry were researchers at the Davis branch of the University of California. As early as the 1930s they had begun mapping the state into microclimates based on growing-season temperatures, and this work emerged as a guide to what areas should be planted in vineyards. In addition, basic research was initiated into all phases of wine making and the results obtained were quickly disseminated into the vineyards and wineries through various degree-granting and training programs. The application of science to wine began in Europe, but in California it reached new heights that set the standard for the late twentieth century.

During the 1960s and 1970s, wine was in a new golden age, and virtually all predictions made at the time foresaw even better days ahead. Many older wine-producing areas of the globe were modernized so that they could market their products abroad. Suddenly wines were appearing from such places as Hungary, Romania, Yugoslavia, and Algeria. However, the most spectacular change took place in Italy, which quickly went from being a producer of wines meant mostly for local consumption to the world's leading exporter. Wine growing also expanded into many non-traditional areas including China, Japan, Kenya, and American states such as Oregon, Texas, and New Mexico. It even returned to England.

The events of the 1980s and early 1990s, however, did not bear out the optimism of the previous two decades, which to a great extent was built around trends in the U.S. market – a market that seemed to be headed in the direction of making Americans into regular wine drinkers on a par with the Europeans. But the course of upward per capita consumption did not proceed as predicted. The adoption of wine as a meal-time beverage has not spread to the population at large; for most people wine retains its image as something either foreign, elitist, or reserved for special occasions. In addition, a growing neoprohibitionist movement centered around the issues of drunk driving and health has exerted its influence on all alcoholic beverages. Information about wine's contribution to cardiovascular functioning has reversed the decline somewhat, but it is too early to tell if that will turn out to be just another health fad.

In Europe, wine consumption has fallen every-

where, and bottled water and soft drinks are becoming preferred by more and more people. In this instance as well, health and safety issues seem to be major reasons for the change in drinking habits.

The impact of declining demand has, to this point, fallen most heavily on lower-priced, so-called jug or country wines. In areas such as the Midi of France, southern Italy, and the central valley of California, supply is greatly in excess of demand. With no signs of change in the offing, the future portent is one of considerable vineyard abandonment in such areas.

By way of contrast, the demand for prestigious high-priced wines has never been greater. Each year they sell out even as their costs escalate. To some extent, this results from wine drinkers preferring quality to quantity – one or two bottles a week instead of four or five. But wine has also become a collectible for the world's affluent. It is not so much a matter of investing in financial terms, but of investing in status. To hold certain types of wine and then to serve them in the appropriate circumstances is to mark one's social standing. The competition among wine producers for this market is intense, with much of it centering on extolling the geographic features that give a wine its special character. The French, for example, argue for the influence of *terroir*, a term that encompasses all the physical characteristics of the growing site. A variation on this theme is increasingly seen in California, where smaller and smaller wine-producing areas and the individual vineyards are indicated on the label. Very recently, Chile has begun exporting larger volumes of wine, stressing that wine makers there attain a unique quality because in that country they grow ungrafted *vinifera* vines. Oregon and Washington State also employ the same selling technique.

To a large extent, the minutiae of locational differences are emphasized in order to counteract a worldwide trend toward uniformity. Only a small number of grapes are considered suitable for making fine wines. These are dominated by Cabernet Sauvignon and Chardonnay, which along with Riesling, Pinot Noir, and two or three others (depending on who is making the list) constitute the so-called noble family of grapes. The standards are set at numerous wine-tasting competitions and by a small group of wine writers whose opinions are enormously influential in shaping the market. As the preferred grapes spread, they provide an underlying sameness to the wines made from them and pose the likelihood of extinction for many long-standing local varieties.

Uniformity results not only from using similar grapes but from similar ways of doing things in the vineyard and winery. Science and technology have created the means for standardizing wine making around the world, and their application has been facilitated by more and more wine coming under the control of large transnational corporations. The regional distinctions of the past are blurring as wine from

places as distant as the Médoc in France, the Maipo Valley in Chile, and the Napa Valley in California all begin to look, smell, and taste alike.

James L. Newman

Bibliography

- Adams, L. 1990. *Wines of America*. Third edition. New York.
- Amerine, M. A. 1983. *Wines, their sensory evaluation*. New York.
- Baxevis, J. J. 1987a. *The wines of Bordeaux and western France*. Totowa, N.J.
- 1987b. *The wines of Champagne, Burgundy, and eastern and southern France*. Totowa, N.J.
- Blij, H. J. de. 1983. *Wine: A geographic appreciation*. Totowa, N.J.
1985. *Wine regions of the Southern Hemisphere*. Totowa, N.J.
- Carter, F. W. 1987. Cracow's wine trade (fourteenth to eighteenth centuries). *Slavonic and East European Review* 65: 537–78.
- Catrell, H., and H. L. Stauffer. 1978. *The hybrids*. Lancaster, Pa.
- Dickenson, J. 1990. Viticultural geography: An introduction to the literature in English. *Journal of Wine Research* 1: 5–24.
- Dion, R. 1959. *Histoire de la vigne et du vin en France*. Paris.
- Grace, V. 1979. *Amphoras and the ancient wine trade*. Princeton, N.J.
- Hyams, E. S. 1965. *Dionysus: A social history of the wine vine*. New York.
- Ilbery, B. W. 1983. The renaissance of viticulture in England and Wales. *Geography* 68: 341–4.
- James, M. K. 1971. *Studies in the medieval wine trade*. Oxford.
- Johnson, H. 1985. *The world atlas of wine*. Third edition. London.
1989. *Vintage: The story of wine*. New York.
- Lachiver, M. 1988. *Vins, vignes et vignerons*. Paris.
- Loubère, L. A. 1978. *The red and the white: A history of wine in France and Italy in the nineteenth century*. Albany, N.Y.
1990. *The wine revolution in France*. Princeton, N.J.
- Muscantine, D., B. Thompson, and M. A. Amerine. 1984. *Book of California wine*. Berkeley, Calif.
- Robinson, J. 1986. *Vines, grapes and wine*. New York.
- Seward, D. 1979. *Monks and wine*. London.
- Stanislawski, D. 1970. *Landscapes of Bacchus*. Austin, Tex.
1975. Dionysus westward: Early religion and the economic geography of wine. *The Geographical Review* 65: 427–44.
- Tehernia, A. 1986. *Le vin de l'Italie romaine: Essai d'histoire économique d'après les amphores*. Rome.
- Teiser, R. 1983. *Winemaking in California: The account in words and pictures of the Golden State's two-century-long adventure with wine*. New York.
- Unwin, T. 1991. *Wine and the vine: A historical geography of viticulture and the wine trade*. London.
- Wagner, P. M. 1956. *American wines and wine-making*. New York.
- Webb, A. D. 1984. The science of making wine. *American Scientist* 72: 360–7.



PART IV

The Nutrients – Deficiencies, Surfeits, and Food-Related Disorders

The discovery of the chief nutrients has been essentially a twentieth-century phenomenon. In 1897, Dutch researcher Christian Eijkman, while investigating beriberi in the Dutch East Indies, showed that a diet of polished rice caused the disease and that the addition of the rice polishings to the diet cured it. Fifteen years later, Polish chemist Casimir Funk proposed that not only beriberi but scurvy, pellagra, and rickets were caused by an absence of a dietary substance he called *vitamine*; and the age of vitamins was under way.

This is not to say that much earlier research did not undergird such twentieth-century breakthroughs. The importance to human health of some minerals, such as iron, had long been at least vaguely recognized, and by 1800, it was understood that blood contained iron; since the eighteenth century, some kind of dietary deficiency had been a periodic suspect as the cause of scurvy; and protein was discovered in the nineteenth century. But in addition to both water- and fat-soluble vitamins, the importance and functions of most of the major minerals and the trace minerals, along with amino acids, were all twentieth-century discoveries, as were the essential fatty acids and the nutritional illness now called protein-energy malnutrition (PEM).

One important consequence of the new knowledge was the near-total eradication of the major deficiency diseases. Pellagra, which had ravaged southern Europe and the southern United States, was found to be associated with niacin deficiency; beriberi, the

scourge of rice-consuming peoples in the Far East, was linked with thiamine deficiency; and scurvy was finally – and definitively – shown to be the result of vitamin C deficiency.

But as rapidly as these illnesses were conquered, new conditions were identified. There was PEM (mostly in the developing world) and anorexia nervosa (mostly in the developed world). Goiter, the result of iodine deficiency, proved to be a problem in both, and Keshan disease was associated with selenium deficiency and the death of children in China.

In addition, faulty nutrition has been implicated in the chronic diseases that are especially prominent in the countries of the West, such as diabetes, heart-related ailments, and cancer. Much of the problem seems to lie in diets significantly higher in sugar, salt, and fats than those of most other peoples – diets of affluence, which among other things produce obesity that can predispose to heart diseases and diabetes. Yet the link between diet and diseases of a chronic nature has often been sufficiently vague as to result, thus far, in more questions than answers. Consequently, many sensible hypotheses abound, and we have included one of these. Our last chapter in Part IV connects a specific nutriment, milk, and lactose tolerance – an ability to absorb milk after childhood, not shared by most of the world's population – with the high rates of coronary artery disease among the people of northern Europe and those descended from northern Europeans in the United States.

IV.A

Vitamins

IV.A.1 Vitamin A

Definitions and Nomenclature

Vitamin A is a fat-soluble substance essential to the health, survival, and reproduction of all vertebrates. As with all vitamins, it is needed in only small amounts in the human diet, about 1 to 1.5 milligrams a day. Vitamin A does not occur in the plant kingdom, but plants supply animals with precursors (or provitamins), such as beta-carotene and other carotene-related compounds (carotenoids), that are converted to vitamin A in the intestinal mucosa of animals and humans. Beta-carotene (and other carotenoids) are abundant in all photosynthesizing parts of plants (green leaves), as well as in yellow and red vegetables. Vitamin A, also known as “retinol,” is itself a precursor of several substances active in the vertebrate organism; these are collectively termed “retinoids.” One retinoid is retinoic acid, an oxidation product of retinol, formed in the liver and other organs and existing in different chemical isomers, such as all-*trans*-retinoic acid and 9-*cis*-retinoic acid, with different functions. Other retinoids are all-*trans*-retinaldehyde and its 11-*cis*-isomer. The latter is active in the retina of the eye, forming the light-sensitive pigment rhodopsin by combination with the protein opsin. In

the liver, retinol is stored in the form of its ester (retinyl palmitate).

Retinoids in the animal organism are generally not found in the free state but are bound to specific proteins. Thus, in the blood, retinol is carried by a retinol-binding protein and, within cells, by an intracellular retinol-binding protein. Retinoic acid and retinaldehyde are carried by specific intracellular binding proteins. When carrying out its hormonal function, retinoic acid combines with another set of proteins, called retinoic acid receptors, located in the cell nucleus. The retinoic acid-receptor complex can then interact with specific genes at sites known as retinoic acid response elements, thereby activating these genes and causing them to stimulate (or repress) the expression of specific proteins or enzymes involved in embryonic development, cell differentiation, metabolism, or growth.

Units

At an earlier time, before pure, crystalline retinol and beta-carotene became available through chemical synthesis, the units used to express vitamin A activity were international units (IU). One IU of vitamin A activity was defined as 0.3 microgram (μg) all-*trans*-retinol (Figure IV.A.1.1a). Because of inefficiencies of

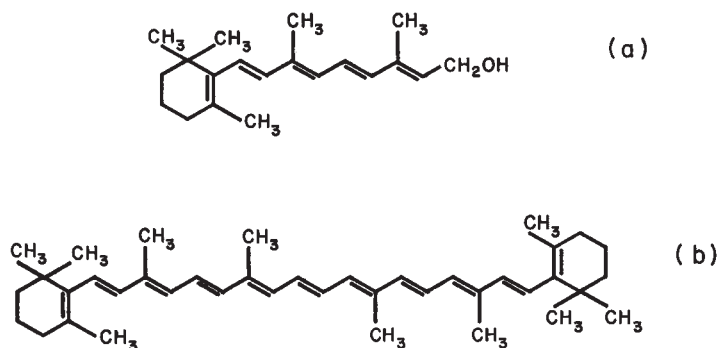


Figure IV.A.1.1. Chemical structure of all-*trans*-retinol (a) and all-*trans*-beta-carotene (b).

absorption and conversion to retinol, 1 IU of beta-carotene was defined as equivalent to 0.6 µg all-*trans*-beta-carotene (Figure IV.A.1.1b). More recently, the accepted unit has become the retinol equivalent (RE), a dietary concept (Underwood 1984). One RE equals 1 µg retinol. Taking into account beta-carotene's average absorbability and conversion to retinol, 1 RE was defined as 6 µg beta-carotene. Of course, it should be noted that absorbability and conversion of dietary beta-carotene can vary greatly with other components of the diet, such as fat or protein.

The Recommended Dietary Intake of vitamin A per day, published by the United States National Research Council, is 500 RE for children, 800 to 1,000 RE for adults, 1,300 RE for pregnant women, and 1,200 RE for lactating women.

History of Vitamin A Deficiency

Night blindness (nyctalopia), meaning defective vision at low illumination, is one of the earliest signs of vitamin A deficiency and was probably the first nutrient-deficiency disease to be clearly recognized. Although the relationship of night blindness to a dietary deficiency of vitamin A was not discovered until the nineteenth century, a cure for an unspecified eye disease, *sharew*, was described in ancient Egyptian texts. Papyrus Ebers (no. 351, c. 1520 B.C.) states: "Another [recipe] for *sharew*. Liver of beef is cooked and squeezed out and placed against it [the eye]. Really excellent." The London Medical Papyrus (no. 35, c. 1300 B.C.) states: "Another recipe [for *sharew*]. Beef liver placed on a fire of straw or emmer or barley and smoked in their [the straw's] smoke; their [the liver's] liquid is squeezed against the eyes" (Wolf 1980: 97).

The word *sharew* apparently refers to a disease of the eyes, and because liver is an especially rich source of vitamin A, and *sharew* was apparently cured by liver, scholars assumed until recently that this disease was caused by vitamin A deficiency and that, therefore, *sharew* meant "night blindness." However, no mention of "night" or "night vision" occurs in any Egyptian source,¹ and we now know that topical application of vitamin A would not be effective in curing night blindness. Much of Egyptian medical practice involved a magic element – the "transfer" of the strength of the ox to the afflicted organ. But the liver would have to be eaten to relieve night blindness. One way that the latter might have occurred has been suggested by A. Sommer, who describes a magical ritual whereby the juice of roasted lamb's liver is applied to the eyes as a cure for night blindness in rural Java today. The liver, however, is not discarded; it is eaten by the patient after the ceremony is finished. Sommer comments that the consumption of the liver is not actually considered part of the ritual. Perhaps the ancient Egyptian ceremony also ended with the patient eating the liver.

The first indisputable recognition of night blind-

ness was by the ancient Greeks, who called it nyctalopia (*nyx*, *nyktos*, "night"; *alaos*, "blind"; *ops*, *opteos*, "eye") and prescribed the oral administration of liver as a cure. The development of Greek medicine was heavily indebted to Egyptian medical theory and practice, and Egyptian medical ideas had an overriding influence on both the Ionian school of medicine and the school of Cos. However, whereas Egyptian physicians considered disease the result of external, supernatural influences, the Greek physicians observed the patient, thought of disease as a condition of the patient, and designed therapy to fit into a rational system, such as the humoral theory in Hippocratic medicine. Hippocrates and his school produced the collection of writings *Corpus Hippocraticum* (Alexandria, c. 300 B.C.), within which the treatise "Concerning Vision" mentioned nyctalopia as an eye disease that usually afflicts children, along with other ailments such as otitis, sore throat, and fever.

The essay recommends this prescription: "Raw beef liver, as large as possible, soaked in honey [presumably to make it more palatable], to be taken once or twice by mouth" (Hirschberg 1899: 100). This is exactly what modern medicine would prescribe if it was not possible to obtain fish-liver extract or vitamin A pills. Today, it is well known that vitamin A deficiency afflicts principally children, predisposing them to infections because of an impaired immune system. Furthermore, fever coupled with infections is now understood to lead to excessive loss of vitamin A from the body stores, aggravating an incipient deficiency state. Thus, one might marvel at the Hippocratic school's acuity of observation and knowledge of effective therapy.

Galen (A.D. 130–200) provided a precise definition of night blindness, as did Oribasius (born c.A.D. 325), a follower of Galen, who defined night blindness this way: "Vision is good during the day and declines at sundown; one cannot distinguish anything any longer at night" (Guggenheim 1981: 267). Although Galen recommended as a cure the "continuous eating of roasted or boiled liver of goats" (Kühn 1821: 803), he also advised topical treatment: The juice of roasted liver should be painted on the eyes – a residue, perhaps, of the Egyptian influence. Much later, around A.D. 540, Aetius of Amida suggested goat liver, fried with salt but not oil, eaten hot. He also suggested that the "juice that runs out of the liver be applied topically to the afflicted eye" (Guggenheim 1981: 267).

The authors of the Assyrian medical texts (c. 700 B.C.) thought night blindness was caused by the rays of the moon and cured by the application of ass's liver to the eyes. Chinese medicine recognized night blindness, and Sun-szu-mo (seventh century A.D.), in his *1000 Golden Remedies*, described a cure by administration of pig's liver (Lee 1967). Arabic medicine, often derived from the translation of Greek medical texts, leaned toward topical application of liver extract (Hunain Ibn Ishaq, ninth century A.D.). So did

a Hebrew treatise from Muslim Spain (Abraham Ibn Ezra, twelfth century), but it added eating liver to the prescription.

The Hippocratic tradition of liver therapy for night blindness survived through the Middle Ages. But with the long sea voyages of discovery that began at the end of the fifteenth century, dietary deficiency diseases, including night blindness, began to appear on board ships. Jacob de Bont (1592–1631), a Dutch physician of the East Indies, described a night blindness that can end in total blindness and prescribed a specific medicament for its cure, namely liver of shark (Lindeboom 1984).

There were many descriptions of night blindness penned by eighteenth- and nineteenth-century Europeans that mentioned that a poor or defective diet was suspected as the cause. In the nineteenth century, for instance, Reginald Heber, Bishop of Calcutta, in *Narrative of a Journey through the Upper Provinces of India: From Calcutta to Bombay* (1824–5), noted:

In our way back through the town a man begged of me, saying that he was blind. On my calling him, however, he came forwards so readily to the torches, and saw, I thought so clearly, that I asked him what he meant by telling me such a lie. He answered that he was night blind ("rat unda"), and I not understanding the phrase, and having been a good deal worried during the day with beggars, for the whole fort is a swarm of nothing else, said peevishly, "darkness is the time for sleep, not for seeing." The people laughed as at a good thing, but I was much mortified afterwards to find that it was an unfeeling retort. The disease of night blindness, that is, of requiring the full light of day to see, is very common, Dr. Smith said, among the lower classes in India, and to some professions of men, such as soldiers, very inconvenient. The Sepoys ascribe it to bad and insufficient food, and it is said to be always most prevalent in scarcity. It seems to be the same disorder to the eyes with which people are afflicted who live on damaged or inferior rice, in itself a food of very little nourishment and probably arises from weakness of the digestive powers. I was grieved to think I had insulted a man who might be in distress, but Dr. Smith comforted me by saying that, even in respect of night blindness, the man was too alert to be much of a sufferer from the cause which he mentioned. (Wolf 1980: 99)

In another nineteenth-century example, during a journey round the globe by an Austrian sailing ship (1857–9), a young ship's doctor, Eduard Schwarz, carried out what appears to have been the first modern rational experiment undertaken to prove or disprove the ancient theory that night blindness could be

cured by eating liver. During the long voyage, several cases of night blindness were observed, and Schwarz reported the following:

[A]t dusk, objects appeared to the afflicted darkened, vague, with indistinct contours and as though moved into the far distance; light-absorbing objects could not be perceived, even close by, and the afflicted walked into the ships' walls or railings. Indoors, everything seemed completely dark to them. We experienced the most convincing proof of the efficacy of beef liver against nightblindness on the return journey, when . . . about 20 nightblind sailors were given ox liver to eat, whereupon they were all permanently cured. (Schwarz 1861: 166)

Interestingly, Schwarz also observed that when the night-blind sailors were kept in complete darkness for three days, their night vision recovered, only to be lost again after a few days in daylight. We can conclude from modern knowledge that during dark adaptation, small amounts of vitamin A reserves remaining in the body reached the retina to reform the depleted visual pigment and were lost again upon exposure to light.

It was also in the nineteenth century that the lesion was noted called xerophthalmia (dryness of the conjunctive and cornea), caused by more severe vitamin A deficiency. V. von Hubbenet described night blindness, associated with "silver scales at the cornea" (Guggenheim 1981), that was caused by a faulty diet but could be treated with beef liver. In 1857, David Livingstone, the explorer of Africa, noting corneal lesions in African natives who subsisted on a diet of coffee, manioc, and meal, wrote: "The eyes became affected as in the case of animals fed pure gluten or starch" (Moore 1957: 4).

At the beginning of the twentieth century, an important contribution was made by M. Mori, who in 1904 described widespread xerophthalmia in Japanese children, including large numbers showing keratomalacia, a softening of the cornea ending in corneal ulceration and necrosis and characteristic, as we now know, of severe vitamin A deficiency. Mori reported that the diet of Japanese children consisted of rice, barley, other cereals, "and other vegetables," and he found that liver, and especially cod-liver oil, were curative. He thus connected the disease with a nutritional cause and a cure (the ingestion of a fat or an oil).

Close on the heels of the discovery of vitamin A (described in the next section) were the observations of C. E. Bloch, who, during World War I, studied children in Denmark afflicted with night blindness and keratomalacia (Guggenheim 1981). As a result of the war, the children had subsisted on a diet of fat-free milk, oatmeal, and barley soup. In an important early experiment, Bloch worked with 32 institutionalized children (1 to 4 years old), dividing them into two groups. One group received whole milk instead of the fat-free milk; the other was given vegetable fat (mar-

garine) but not whole milk. The latter showed 8 cases of corneal xerosis, the former remained healthy. All the xerosis cases were cured rapidly with cod-liver oil. Bloch correctly surmised that whole milk, cream, butter, eggs, and cod-liver oil contained lipid substances protecting against corneal and conjunctival xerosis.

History of the Discovery of Vitamin A

The first animal experiments with a vitamin A-deficient diet were those of F. Magendie in 1816 (McCollum 1957). He fed dogs a diet of wheat gluten, or starch, or sugar, or olive oil and described the corneal ulcers that developed (as we now know) because of a lack of vitamin A.

The effects of vitamin A deficiency on growth were first described by G. Lunin in 1881. He found that mice could not survive on a diet of pure casein, fat, sucrose, minerals, and water, but they lived and grew normally with whole dried milk. The outcome of similar experiments led F. G. Hopkins (1906) to postulate “minimal qualitative factors” in the diet necessary for growth and survival (Moore 1957).

An interesting and exciting account of the beginnings of experimental nutrition, which ultimately led to the discovery of vitamin A, was written by E. V. McCollum (1964). At the time McCollum began his research (1907), chemical analysis of foods and animal feed was well advanced. He fed cows rations of hay with wheat, or oats, or yellow maize, of identical composition as far as protein, carbohydrate, fat, and minerals were concerned.

The wheat-fed cows did not thrive but, rather, became blind and gave premature birth to dead calves. The oat-fed cows fared somewhat better, but the maize-fed cows were in excellent condition, produced vigorous calves, and had no miscarriages. McCollum spent four years in trying to determine the reasons for these remarkable results. He addressed questions such as: Did the wheat contain a toxic substance, and what was lacking in wheat that was present in yellow maize? He then conceived the important idea that nutrition experiments could be done much faster and better with small animals, such as mice or rats. They eat much less than larger animals, they can be given purified diets, they reproduce rapidly, and many can be housed at one time (McCollum 1964).

Raising rats, however, was not a popular idea at the Wisconsin College of Agriculture, where McCollum worked, because, as he was told by the dean, “rats were enemies of the farmer.” Thus, he surreptitiously started a rat colony in a basement of the Agricultural Hall. In 1915, he found that whereas young rats on a diet of pure protein, pure milk sugar, minerals, and lard (or olive oil) failed to grow, the addition to their diet of butterfat or an ether extract of egg yolk restored their health. Soon after, he determined that adding an ether extract of alfalfa leaves or, even bet-

ter, of liver or kidney, greatly improved the diet's growth-promoting quality. Clearly, he had found a fat-soluble factor that promoted growth in rats.

He saponified butterfat, extracted the unsaponifiable mixture into ether, and added the extract to olive oil. Now the olive oil, with pure protein, sugar, and minerals, supported growth. This experiment, and subsequent ones, proved that butterfat contained a fat-soluble growth factor (other than fat itself) that could be transferred from one fat to another. McCollum ultimately showed that this factor was essential for growth and survival. This was the discovery of vitamin A, then named “fat-soluble factor A” to distinguish it from other accessory dietary factors that were grouped under the term “water-soluble B.” Simultaneously, biochemists T. B. Osborne and L. B. Mendel made the same discovery; they published their findings five months later (1915).

In 1919, it was noted by another investigator, H. Steenbock, that active growth-promoting extracts, such as those made from butter, egg yolk, or certain plants (carrots), were yellow, whereas extracts from lard, inactive in growth promotion, were white (Table IV.A.1.1). He realized, however, that active fat-soluble factors existed in white extracts from liver or kidney. Steenbock and P. W. Boutwell thereupon proposed the hypothesis in 1920 (it was not fully confirmed for another 10 years) that fat-soluble factor A was associated with a yellow pigment (now known to be beta-carotene) and suggested that the yellow pigment could be converted to an active colorless form (vitamin A or retinol).

In the meantime, L. S. Palmer (in 1919) threw the subject into confusion by feeding chickens white maize, skimmed milk, and bone meal, a diet free of yellow pigments. When the birds stopped growing and began to fail, he supplemented their diet with a small amount of pork liver, containing no yellow pigment, and found that though the birds had received no yellow pigment, they grew normally and laid eggs. The egg yolks were colorless, but the eggs, when hatched, produced normal chicks.

Table IV.A.1.1. Association of vitamin A potency with yellow color in food

Active (yellow) (white)	Inactive, or low potency
Butter	Casein
Egg yolk	Egg white
Cod-liver oil	Lard
Yellow maize	White maize
Carrot	Parsnip
Sweet potato	Ordinary potato
Red palm oil	Palm kernel oil
Outer green leaves of cabbage	Inner white leaves of cabbage
Tomato	Onion
Apricot	Apple

Source: From Moore (1957).

It was not until 1929 that T. Moore resolved the conflict by proving that the yellow pigment extracted from plant sources, butterfat, or egg yolk, and purified as carotene, was converted to the active factor in the animal body. He fed pure, crystalline carotene (yellow) to young rats and observed the accumulation of retinol (colorless) in their livers. He correctly concluded that yellow carotene is the provitamin or precursor of colorless vitamin A (retinol).

Discovery of the Functions of Vitamin A

In the 1930s, the isolation and synthesis of both carotene (now more precisely defined as beta-carotene) and retinol were accomplished, along with an understanding of their chemical structures. The enzymatic conversion of beta-carotene to retinol was later shown to take place in the intestinal mucosa of animals and humans. The requirements and recommended daily allowances of beta-carotene and retinol for animals and humans were also determined.

Early research into the function of retinol was based on physiological and biochemical observations of the responses of experimental animals to diets lacking in beta-carotene or retinol. When weanling rats were fed a vitamin A-deficient diet, their liver stores of retinyl esters became exhausted after about four weeks, and their growth began to decline after five weeks. They then became night blind, and their corneas began to show lesions and to ulcerate after about six weeks on the deficient diet. After this they lost weight precipitously, became sick, and died. This simple feeding experiment highlighted three of the most important physiological functions of vitamin A: It promotes vision, normal differentiation of epithelia (such as the cornea), and growth. Vitamin A deficiency detracts from these functions and, in addition, leads to abnormal bone growth, inhibition of spermatogenesis in male animals, and resorption of the fetus in pregnant females.

The visual function of vitamin A was elucidated in great detail and with exceptional brilliance (culminating in a Nobel Prize) by G. Wald. As we have already noted, night blindness (defective vision at low illumination) was recognized in the eighteenth century as an early symptom of a dietary deficiency; later research indicated that night blindness, and its associated lesions, resulted specifically from a deficiency of vitamin A. It was, therefore, fitting that the function of vitamin A in vision was the first to be defined with a precise mechanism at the molecular level. As Moore (1957: 263) commented: "It may be an inspiring thought . . . that Man's knowledge of the existence of the stars and the vast universe which appears in the heavens each night, comes in the first place from the stimulation by the light rays of delicately poised molecules of vitamin A."

As long ago as 1878, W. Kühne recognized that retinas from dark-adapted frogs, purple in color, changed

to yellow when exposed to light. The connection with vitamin A was made in 1925 by L. S. Fridericia and E. Holm, who observed that vitamin A-deficient, light-adapted rats, when put into the dark, formed visual purple at a slower rate than normal rats. Then, in 1929, Holm observed the presence of vitamin A in the retina. However, it was Wald, beginning in 1935, who found that the visual purple of the retina, called rhodopsin, consists of a protein, opsin, combined with "retinene" (later shown by R. A. Morton to be retinaldehyde). Wald and his co-workers (Wald 1968: 230-9) described the visual cycle thus: The 11-*cis*-isomer of retinaldehyde, while attached to opsin in rhodopsin, is isomerized to the all-*trans* form by light; this is the event that triggers the nerve impulse to the brain as perception of light. The all-*trans*-retinaldehyde is released from the opsin and reduced to all-*trans*-retinol. This, in the dark, is isomerized to 11-*cis*-retinol, oxidized to 11-*cis*-retinaldehyde, which recombines with opsin to re-form rhodopsin (dark adaptation), thus completing the cycle.

Elucidation of the metabolic function of vitamin A lagged behind the recognition of its role in vision. The observations of S. B. Wolbach and P. R. Howe (1925) defined the epithelial lesions resulting from vitamin A deficiency, particularly the keratinization of epithelial cells, especially of the cornea, but also of the respiratory, intestinal, and genitourinary tracts. It seemed that vitamin A influenced the differentiation of epithelial cells from the normal simple and pseudostratified phenotype to squamous, metaplastic lesions, starting focally and, ultimately, spreading throughout the epithelium.

Apart from the function of maintaining normal epithelial differentiation, the pathology of vitamin A deficiency also revealed a function of vitamin A in bone growth, shown by the thickened bones in growing, vitamin A-deficient animals. In addition, the deficiency is manifest in reproductive functions; in the male, by cessation of spermatogenesis and atrophy of the testes; in the female, by a defective uterus, placental necrosis, and, ultimately, resorption of the fetus. Indeed, as long ago as 1933, the teratogenic effect of only marginal vitamin A deficiency, which resulted in ocular and genitourinary defects in pig embryos, underlined the necessity of vitamin A to support normal fetal development.

In the 1950s and 1960s, the metabolism of vitamin A was delineated by J. A. Olson and others, and recognition of the importance of its transport proteins followed. This discovery was made by D. S. Goodman with regard to the plasma retinol-binding protein, and by F. Chytil and D. E. Ong with respect to the intracellular-binding proteins. K. R. Norum and R. Blomhoff and others established the basic facts of absorption and liver-storage mechanisms for retinol.

In the 1970s and 1980s, the discovery by U. Safiotti of an anticarcinogenic action of vitamin A (although still a matter of some controversy) led to an

outburst of research effort. In animal experiments, retinol, and particularly its metabolite, retinoic acid, have been demonstrated indisputably to have anticarcinogenic properties. But as applied to human cancer patients, a search for similar activity was disappointing, although a number of isomeric forms and synthetic chemical derivatives of retinoic acid were found to be active against some forms of cancer. For example, acute promyelocytic leukemia has been successfully treated with all-*trans*-retinoic acid; certain skin and head-and-neck cancers have yielded to 13-*cis*-retinoic acid, and retinoic acid and some of its synthetic chemical derivatives have exhibited spectacular cures in a number of skin diseases.

A new era of vitamin A research dawned in 1987, when P. Chambon in Strasbourg, France, and R. M. Evans in San Diego, California, and their respective co-workers, simultaneously discovered the retinoic acid receptors in cell nuclei (Mangelsdorf, Umesono, and Evans 1994). These receptor proteins can bind retinoic acid and, when thus liganded, can activate a number of specific genes to stimulate the cells either to produce specific proteins (enzymes) or to inhibit the activity of other enzymes. In this way, one can arrive at an explanation at the molecular level of the many metabolic functions of vitamin A with regard to embryonic development, differentiation, and growth. This type of action – gene activation – establishes vitamin A (in the form of its metabolite, retinoic acid) as a hormone, similar to the steroid hormones and the thyroid hormone. As was predicted in 1970: “[T]here is no *a priori* reason why a hormone should have to be made by the animal itself. It is quite conceivable that a hormone can be taken in the diet [as vitamin A], stored in the liver, and secreted into the bloodstream when needed. The liver then acts as the endocrine organ” (Wolf and De Luca 1970: 257).

Clinical Manifestations of Vitamin A Deficiency

Vitamin A deficiency diseases appear primarily in the young because, like other mammals, humans are born with low liver reserves. The latter probably results from a placental transport of retinol, which, although generally sufficient for normal development, is not sufficient for building any surplus. After birth, colostrum (and later breast milk) provides the needed vitamin, so long as the mother's intake of vitamin A is adequate. If, after weaning, vitamin A stores have not been built up, and malnutrition plagues the child, deficiency signs will appear. Other causes for vitamin A deficiency are protein malnutrition, when the retinol transport system in blood does not function normally, and chronic malabsorption diseases such as cystic fibrosis.

The first sign of deficiency, as already discussed, is night blindness. With increasing severity of the deficiency, the following ocular signs are exhibited: con-

junctival xerosis; Bitot's spots (white, foamy spots on the conjunctival surface, consisting of sloughed-off cells); corneal xerosis; and corneal ulceration (also called keratomalacia), which often ends in irreversible blindness, a direct consequence of the liquefaction of the corneal stroma. Recently, increased morbidity and mortality, frequently connected with diarrhea and respiratory diseases, has been observed in children with marginal vitamin A intake and no eye lesions (see next section), no doubt caused by a compromised immune system (Underwood 1994). It has also been noted that measles in marginally depleted children can precipitate severe vitamin A deficiency diseases. Figure IV.A.1.2 presents a diagrammatic representation of the deficiency signs.

In adults, vitamin A deficiency gives rise, apart from night blindness, to skin lesions (hyperkeratosis of hair follicles). Moreover, experimentally induced vitamin A deficiency in adult human volunteers caused raised cerebrospinal fluid pressure, loss of the senses of taste and smell, and abnormalities in vestibular function (Wolf 1980: 146–50).

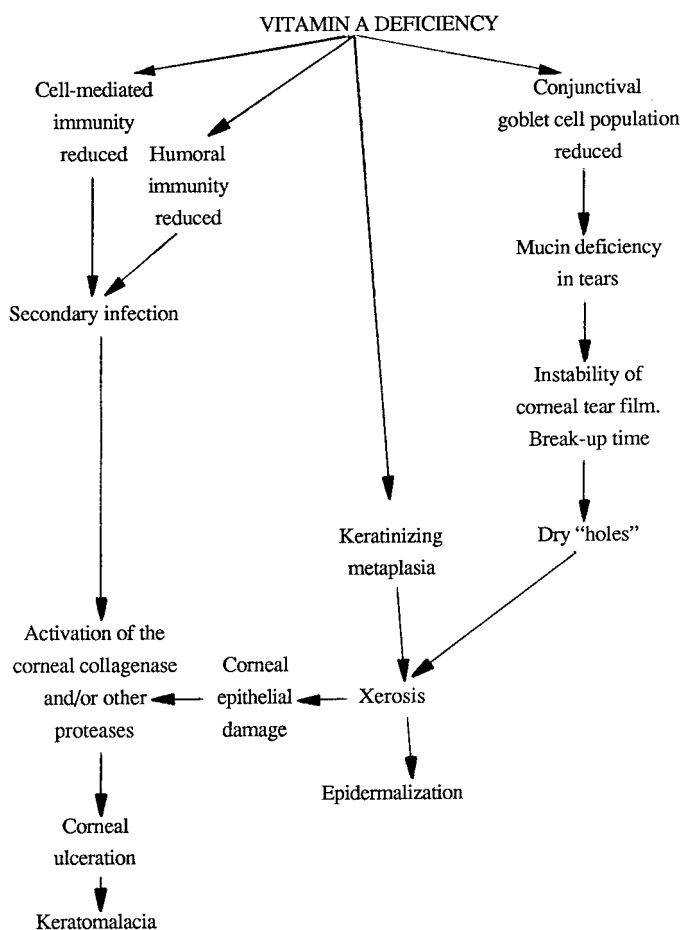
Dangers of Vitamin A Excess

If too little vitamin A is harmful, so also is too much, and emphasis needs to be placed on the dangers of hypervitaminosis A. In technically developed countries, many cases of vitamin A intoxication by ingestion of large doses of retinol or retinyl ester supplements (25,000 IU or more over a prolonged period, i.e., one month or longer) have been reported. It is not difficult to understand the reasoning behind this behavior (“if a little is good for me, a lot must be better”), and such reasoning is encouraged by popular literature, media exaggerations, a vague idea that vitamins generally are beneficial, and incomplete reports of the anticancer effects of vitamin A.

Unfortunately, such reasoning with vitamin A can lead to vitamin A intoxication, with symptoms such as skin rash, desquamation and pruritus of the skin, hair loss, weakness and fatigue, liver damage, bone and joint pain (associated with osteopenia of vertebrae and calcification of spinal ligaments and periosteal tissues in extreme cases), anorexia, and an increase in cerebrospinal fluid pressure, sometimes prompting a misdiagnosis of brain tumor. Hypervitaminosis A is rapidly reversible after the patient stops taking the vitamins. (It should also be emphasized, however, that for metabolic reasons, the provitamin A, beta-carotene, is nontoxic at any dose.)

In addition, recent oral use of an oxidation product of retinol (13-*cis*-retinoic acid) as a drug for the very successful treatment of acne has led to cases of fetal malformations, and there is a risk of abnormalities in the babies if the drug is taken during pregnancy, even at the very earliest stages after conception. Extreme caution is, therefore, indicated to avoid this drug if there is a possibility of pregnancy.

Figure IV.A.1.2. A postulated mechanism for the pathogenesis of keratomalacia in vitamin A deficiency
(From unpublished documents, Nutrition Unit, World Health Organization, Geneva, Switzerland, 1994. With permission)



Geographic Distribution and Epidemiology of Vitamin A Deficiency

Vitamin A deficiency clusters in periequatorial areas and is closely associated with areas of poverty and malnutrition. Estimates suggest that about 350,000 preschool children become blind every year because of severe vitamin A deficiency; of these children, 60 percent die within a year (Underwood 1994).

Recent epidemiological surveys have demonstrated that mortality rates of preschool children receiving low or marginal dietary vitamin A, *yet not showing ocular signs of vitamin A deficiency*, could be lowered by 20 to 50 percent through vitamin A supplementation. Meta-analysis revealed a protective effect in lowering mortality by 23 percent in eight different studies done in Indonesia, Ghana, Nepal, India, and Sudan. From these data it is estimated that worldwide, 13 to 14 million children suffer from mild deficiency (no night blindness) with consequent increased risk of childhood infections (measles, respiratory infections, diarrhea). Vitamin A supplementation lowered the risk of death from measles in areas

where subclinical vitamin A deficiency was prevalent. The cause of the protective effect of vitamin A is not known, but vitamin A depletion probably compromises the immune system.

Clearly, vitamin A deficiency and depletion are very serious public health problems. The World Health Organization actively monitors and obtains data on the worldwide occurrence of clinical and subclinical vitamin A deficiency and recommends ways to improve the situation. Intervention programs to eliminate the deficiency have been successful in Indonesia and are being implemented in India, Bangladesh, and Tanzania (Underwood 1994). Such programs involve food fortification, distribution of supplements, dietary diversification, and control of infectious diseases. The latter is important because there is a synergistic relationship between infectious diseases and vitamin A deficiency: Subclinical deficiency makes children prone to infectious diseases, and such diseases, in turn, can increase the need for the vitamin and precipitate or worsen a deficiency.

It is ironic and tragic that vitamin A deficiency occurs mostly in periequatorial regions where the

Table IV.A.1.2. Countries categorized by degree of health; importance of vitamin A deficiency by WHO region (From information available to WHO in May 1994)

WHO region	Clinical	Subclinical			No data: problem likely
		Severe ^a	Moderate ^b	Mild ^c	
Africa	Benin*	Cape Verde*	Angola*	Liberia*	Botswana
	Murkina Faso	Congo	Guinea*	Madagascar	Algeria
	Cameroon	Gambia	Namibia		Burundi
	Chad	South Africa	Sierra Leone*		Central African Republic
	Ethiopia	Zambia	Zaire*		Côte d'Ivoire
	Ghana		Zimbabwe*		Equatorial Guinea
	Kenya				Gabon
	Malawi				Guinea Bissau
	Mali				Lesotho
	Mauritania				São Tome & Principe
	Mozambique				Swaziland
	Niger				
	Nigeria*				
	Rwanda*				
	Senegal				
	Togo				
	Uganda				
	United Republic of Tanzania				
Americas	Brazil	Mexico	Bolivia	Belize	Cuba
	Dominican Republic	Nicaragua	Ecuador	Colombia	
	El Salvador*	Peru	Honduras*	Panama	
	Guatemala				
	Haiti				
South-East Asia	Bangladesh		Bhutan		Maldives
	India		Indonesia		
	Myanmar		Sri Lanka*		
	Nepal		Thailand		
Eastern Mediterranean	Afghanistan*	Pakistan	Djibouti	Egypt	Iraq
	Somalia		Oman*	Iran	Jordan
	Sudan			Saudi Arabia	Lebanon
	Yemen				Syria
Western Pacific	Cambodia	Papua New Guinea*	China	Malaysia	
	Federal States of Micronesia		Solomon Is.*	N. Mariana Islands*	
	Kiribati				
	Lao People's Dem. Rep.				
	Marshall Is.				
	Philippines				
	Vietnam				

^aUrgently important public health problem in the whole, or in a substantial part, of the country based on serum retinol levels.

^bModerately important public health problem in the whole, or in a substantial part, of the country (subclinical biochemical and/or dietary evidence).

^cSporadic occurrence or high risk of a public health problem in part of the country based on dietary and ecological risk factors.

*Insufficient data to be certain of category.

Source: Indicators for assessing vitamin A deficiency and their application in monitoring and evaluating intervention programs and unpublished documents, Nutrition Unit, World Health Organization (Geneva: 1994). With permission.

populations are surrounded by sources of the provitamin, beta-carotene, in the form of green leaves. It is thought that in rice-dependent countries (India, Bangladesh, Indonesia, and the Philippines, for example), vitamin A depletion is caused by the practice of providing weaned children a mostly rice diet, which lacks both vitamin A and beta-carotene. Such ignorance of good nutrition practices could be combated by nutrition education that would lead to the preparation of beta-carotene-rich foods for children.

The principal causes of the deficiency, however, are poverty and infectious diseases. To quote B. A. Underwood: "Rural areas in the affected regions are generally characterized by economic and social underdevelopment, including the quality and distribution of health services, sanitary facilities and potable water. These factors, characteristic of deprivation, are associated with high morbidity and consequently increased nutrition needs." Underwood stresses the "... environmental factors that can alter requirements

and increase risks of hypovitaminosis A. Such risks are clustered in households and neighborhoods where poverty, poor personal and environmental sanitation, functional illiteracy, social isolation, inadequate access to health and social services prevail" (Underwood 1984: 355).

Vitamin A deficiency is generally found in regions of the globe where protein-energy malnutrition is also prevalent. Animal experiments have shown that a low-protein diet reduces the intestinal enzymes necessary for conversion of beta-carotene to retinol. Thus, a double danger threatens children in such regions: a short supply of beta-carotene *and* protein depletion that may reduce the conversion of available beta-carotene to retinol.

Awareness of this huge problem has finally galvanized governments, and plans were put in motion by the World Health Organization to eliminate vitamin A deficiency by the year 2000.

Vitamin A and Carotenoids in Food

Preformed vitamin A (retinol) is found only in animal foods, with the highest levels in mammalian, fowl, and fish livers, and in fish-liver oil, but with great seasonal variation. The next highest levels are in kidneys and fatty fish (e.g., salmon), whereas a third tier consists of milk, butter, and eggs, again with seasonal variation.

Carotenoids (beta-carotene and related compounds) are precursors of vitamin A that occur in green leaves and yellow and red vegetables and constitute the main source of vitamin A in the human diet. In decreasing order of beta-carotene content, these sources include parsley, sweet potato, broccoli, lettuce, tomato, and cabbage. Smaller, but significant, amounts are in legumes such as chickpeas and green and black grams. Yellow maize also contains beta-carotene. Fowl that eat grains and seeds absorb and accumulate carotenoids in their fat, so that their flesh can also be a source of this provitamin.

A special case as a source of beta-carotene is red palm oil, a vegetable oil used for cooking in several areas of Africa, Southeast Asia, and South America. It contains 0.6 to 1 RE per gram of oil and, in these areas, represents an important source of the provitamin (Underwood 1984: 290). The oil is extracted from the fleshy mesocarp of the palm nut and is prepared by boiling the fruit in water, cooling, and removing the oily layer. Unfortunately, substantial losses of beta-carotene can occur during cooking and storage.

Foods that lack vitamin A or carotenoids are wheat, oats, rice, potatoes, cassava, manioc, onions, nuts, sugar, olive oil, pork, and lard. In light of this lack, it is of interest to consider traditional diets of children in countries where vitamin A deficiency is a widespread public health problem. J. M. May and D. L. McLellan (1970), for example, have described the foods given to children in rural Ethiopia after weaning (at about 18 months of age). These included pancakes

(*injera* - made from a local low-protein cereal grain called *toff* [tef], fermented and wet-cooked on a large griddle over an open fire) and a stew (*wot* - made of legumes, barley, onions, and garlic). The *wot* may contain some meat, although two-thirds of rural Ethiopians never eat meat. The children of a family are especially disadvantaged in terms of nutrition because the order of feeding is first the father, then guests, male persons, mother, and, last, children. Often, by the time the children eat, the *wot*, containing small amounts of carotenoids, is gone, and only *injera* is left.

Another example is drawn from Ghana, where the principal diet in rural areas consists of either fermented maize flour (*keuke*) made into a dough and cooked, or a soup or porridge of millet (*tuwonsafe*). In tropical forest regions, sun-dried cassava is ground to a flour (*konkote*) and boiled to a puree. Any oils in the diet are from groundnuts. Once again, fathers eat first, then other males, with women and children last. Moreover, there are taboos against women consuming eggs in some parts of Ghana. Clearly, with such a diet and such cultural habits, vitamin A deficiency is likely to be widespread.

A final example is found in India, where a public health problem exists with respect to vitamin A deficiency: The poor eat principally parboiled rice, or millet or wheat flours mixed with water and cooked as pancakes (*chapattis*). Cooking oil is from seeds; legumes are cooked with spices, and fat and vegetable intake is low (1 percent of the diet). Children are weaned at 2 years, then given boiled milk, diluted with water, along with rice or *chapattis*. No wonder that vitamin A deficiency diseases are prevalent among the poor.

Future Prospects

Basic Biology

Since the discovery, in 1987, of the mechanism of gene activation by retinoic acid, establishing vitamin A as a hormone, progress in the understanding of its basic biology has been breathtakingly rapid. The pleiotropic action of vitamin A (i.e., the multiple physiological functions of this vitamin) in development, growth, and differentiation, and the legion of enzymes affected by it, has been a puzzle since the early work of Wolbach and Howe (1925). This puzzle is now in the process of being solved.

The regulation and control of growth factors, hormones, enzymes, and homeobox genes (i.e., genes controlling embryonic development) by retinoic acid can now be explained on a rational basis because of the discovery of nuclear retinoic acid receptors (RARs), proteins that bind metabolites of retinol (all-*trans*-retinoic acid and 9-*cis*-retinoic acid) and, in a variety of combinations, interact with the DNA in the promoter regions of a large number of genes. These genes are then activated to express specific mRNAs (messenger RNAs) and proteins that can regulate and control development, growth, and differentiation.

Researchers are sequencing the relevant genes and tracing their evolutionary development. They are also in the process of elucidating the genes' interaction with RARs and the functions of the specific proteins expressed (enzymes, hormones, growth factors). Thus, the fields of biochemistry, endocrinology, genetics, embryology, evolution, and oncology will be greatly expanded, all on the basis of our recognition of the nuclear RARs. It has now become clear that the immune system is also influenced by vitamin A, and important advances in research on the action of vitamin A on the immune response can be expected.

Cancer and Other Diseases

Acute promyelocytic leukemia and skin cancer have been successfully treated with vitamin A derivatives. On the basis of experimental work, one can expect other forms of cancer to yield to efforts at treatment by retinoids. Chemoprevention treatment by retinoids of persons at high risk (e.g., recurrent squamous carcinoma of the head and neck) will certainly expand. Skin diseases (acne, psoriasis, and others) have been successfully treated with retinoids, and the possibility exists that rheumatoid arthritis and other inflammatory diseases may ultimately yield to retinoid treatment.

Public Health

Future prospects for the alleviation of the horrendous tragedy of preventable childhood blindness and children's infectious diseases by vitamin A or beta-carotene now lie in the hands of governments, international organizations, and nongovernmental agencies. The problem is not one of science but of economics, education, and, ultimately, politics. The science is merely concerned with methods of determining in which population groups the children are at risk of vitamin A deficiency and the severity of the deficiency. This assessment can be done by blood tests, but simpler methods are being sought. Once the deficient population group has been ascertained, it is necessary to direct efforts for a cure in such a way that the vitamin reaches the children in need. This implementation involves improving the children's overall diet and sanitation, removing parasites, educating the mothers, diversifying the food supply, fortifying foods, and supplementing with vitamins. Considering the millions of children and the multitude of countries affected (see Table IV.A.1.2), each with its own government, bureaucracy, and dietary and cultural traditions, the elimination of vitamin A depletion and deficiency will be a daunting task.

George Wolf

The author thanks Dr. Barbara A. Underwood of the WHO Nutrition Unit, Geneva, Switzerland, for making available and giving permission to use material from unpublished documents from the World Health Organization.

Notes

1. The author is grateful for this information to Dr. John F. Nunn, Northwood, Middlesex, England.

Bibliography

- Guggenheim, K. Y. 1981. *Nutrition and nutritional diseases, the evolution of concepts*. Boston, Mass.
- Hirschberg, J. 1899. Geschichte der Augenheilkunde. In *Handbuch der gesamten Augenheilkunde*. Second edition, ed. T. Saemische, Vol. 12, Part 2, 98–106. Leipzig.
- Kühn, C. G. 1821. *Galenii omnia opera*, Vol. 19, 802–3. Leipzig.
- Lee, T. 1967. Nyctalopia. In *Diseases in antiquity*, ed. D. B. Fothwell and A. T. Sandison, 418–19. Springfield, Ill.
- Lindeboom, G. A. 1984. Geschichtliche Meilensteine in der Behandlung der Nachtblindheit. *Clio Medica* 19: 40–9.
- Mangelsdorf, D. J., K. Umeson, and R. M. Evans. 1994. The retinoid receptors. In *The retinoids*, ed. M. B. Sporn, A. B. Roberts, and D. S. Goodman, 319–50. New York.
- May, J. M., and D. L. McLellan. 1970. *The ecology of malnutrition*. New York.
- McCollum, E. V. 1957. *A history of nutrition*. Boston, Mass.
1964. *From Kansas farm boy to scientist*. Lawrence, Kans.
- Moore, T. 1957. *Vitamin A*. New York.
- Schwarz, E. 1861. *Reise der oesterreichischen Fregatte Novara um die Erde*. Vienna.
- Sommer, A. 1978. Cure for nightblindness. *American Journal of Clinical Nutrition* 31: 1489.
- Sporn, M. B., A. B. Roberts, and D. S. Goodman, 1994. eds. *The retinoids*. Second edition. New York.
- Underwood, B. A. 1984. Vitamin A in animal and human nutrition. In *The retinoids*, Vol. 1, first edition, ed. M. B. Sporn, A. B. Roberts, and D. S. Goodman, 282–392. New York.
1994. Vitamin A in human nutrition: Public health considerations. In *The retinoids*. Vol. 1, second edition, ed. M. B. Sporn, A. B. Roberts, and D. S. Goodman, 211–28. New York.
- Wald, E. 1968. Molecular basis of visual excitation. *Science* 162: 230–9.
- Wolbach, S. B., and P. R. Howe. 1925. Tissue changes following deprivation of fat-soluble vitamin A. *Journal of Experimental Medicine* 47: 753–77.
- Wolf, G. 1980. Vitamin A. In *Human nutrition, a comprehensive treatise*, ed. R. B. Alfin-Slater and D. Kritchevsky, 96–100. New York.
- Wolf, G., and L. M. De Luca. 1970. Recent studies on some metabolic functions of vitamin A. In *Fat-soluble vitamin symposium*, ed. H. F. DeLuca and J. W. Suttie, 257–65. Madison, Wis.

IV.A.2 🐉 Vitamin B Complex: Thiamine, Riboflavin, Niacin, Pantothenic Acid, Pyridoxine, Cobalamin, Folic Acid

The history of the discovery of the B vitamins includes both the recognition that particular diseases can result from dietary inadequacies and the subse-

quent isolation of specific nutrients from foods that have been found to prevent and to cure those diseases. Most of these dietary deficiencies were first recognized in humans, but in certain instances, the deficiency was produced by feeding restricted diets to experimental animals.

After each of the B vitamins was isolated, extensive biochemical and physiological studies were conducted to define their specific functions. States of B vitamin dependency were also discovered in which the need for a particular vitamin exceeded the physiological level. These vitamin dependencies were found either to have a genetic basis or to be drug-induced.

Moreover, recognition that certain synthetic compounds bearing close chemical resemblance to B vitamins could block the activity of naturally occurring vitamins has led to a better understanding of vitaminic functions and has provided us with certain drugs that are used in cancer chemotherapy and in the treatment of infections and inflammatory diseases.

Vitamin B research, as well as clinical and public health information, is summarized here to emphasize some significant advances in our knowledge of these compounds. But first a note on nomenclature seems appropriate.

B Vitamin Nomenclature

The confusing nomenclature of the B vitamins can only be understood historically. In 1915 E. V. McCollum and M. Davis concluded that young rats needed two unknown "growth factors," one fat-soluble and the other water-soluble, and for convenience they called them "factors A and B," respectively. Casimir Funk had already coined the term "vitamine" in the belief that any unknown factors would prove to have the chemical structure of "amines." J. C. Drummond then suggested that the term be modified to "vitamin," which had no chemical significance, and he called the antiscurvy factor "vitamin C."

Further research showed that "factor B" was made up of at least two components, one heat-labile and the other more stable; these were termed vitamins B₁ and B₂ respectively. Vitamin B₁ was later shown to be the anti-beriberi vitamin, known as thiamine. The B₂ factor was again found to be complex. Among other things, it included riboflavin, which was first named "vitamin G" in honor of Joseph Goldberger, who had made classical studies of pellagra in the United States. The designation vitamin B₃ was given to a compound that seems to have been pantothenic acid but was sometimes incorrectly used to mean niacin. Biotin was first called "vitamin H" by workers in Germany who had shown that its deficiency in rats resulted in abnormalities of the skin (in German *Haut*), and folic acid, at one time, was termed "vitamin M" because it was first shown to be needed by monkeys. Today, however, these three letters, namely, G, H, and M, are no longer used, and the vitamins are known by the

names riboflavin, biotin, and folic acid and classified as parts of the vitamin B complex.

Other workers claimed higher numbers in the B series when they believed that they had discovered new growth factors. But many of these assertions seem to have been erroneous claims and only two "B" numberings are still in use: B₆ for pyridoxine and closely related compounds, and B₁₂ for cobalamin and its derivatives. The coding is convenient because in each case it covers more than one precise chemical compound.

In recent years, some entrepreneurs engaged in selling materials that have been claimed to cure cancer and other diseases have named their products as vitamins in attempts to portray them as essentially natural materials, and so to escape the strict regulations applying to drugs. An example is B₁₇, or "laetrile," which is not recognized by orthodox nutritionists.

Thiamine

Use of polished rice as a dietary staple has long been known to lead to the development of a disease known as beriberi, which was first described as affecting Japanese sailors as well as prisoners in the Dutch East Indies. Signs of beriberi include paralysis due to polyneuritis and congestive heart failure. Christian Eijkman, a medical officer in Java, was the first to show (1890) that a paralytic illness resembling beriberi could be produced in chickens by feeding them polished rice (Jansen 1950). He, and his successor Gerrit Grijns, also demonstrated that the polyneuritis of beriberi could be cured by feeding rice bran.

In 1912, Funk extracted a substance from rice bran that he believed was the anti-beriberi substance and that he characterized chemically as being an amine. He then coined the term "vitamine" as a contraction of "vital amine." With the discovery of other organic compounds that were "vital" but not amines, the term "vitamin" was used instead. The anti-beriberi vitamin was named thiamine. The biologically active coenzyme form of thiamine is required for the oxidative decarboxylation of pyruvic acid. Thiamine is also required in the metabolism of glucose and many other enzyme-catalyzed reactions.

It has been shown that alcoholics may develop an acute thiamine deficiency if they drink heavily and concurrently go without food. Their condition of acute thiamine deficiency, which is manifested by a confusional state and paralysis of certain eye muscles, is designated as Wernicke's encephalopathy. It responds to high doses of intravenously administered thiamine. If this confusional state is not recognized and treated with thiamine, a chronic organic brain disorder develops, associated with dementia. Those who have this disorder may require institutionalization. Korsakoff's psychosis is believed to be a chronic form of Wernicke's encephalopathy, and the health costs of caring for persons with this disease have led to the fortifica-

tion of beers with thiamine in an experimental public health intervention in Australia (Yellowlees 1986).

Nicotinic Acid, or Niacin

Pellagra is a disease that we now recognize as being caused by a deficiency in nicotinic acid (niacin). It is characterized by development of a dermatitis, which appears on areas of the body exposed to strong sunlight; by diarrhea, associated with malabsorption of nutrients; and by a severe confusional state. This disease was first described by Gaspar Casal, who wrote of the diet of Spanish peasants with pellagra or *mal de la rosa*, as it was then called (the book was published posthumously in 1762).

Because the Spanish peasants who had this disease subsisted on cornmeal, and because corn, particularly moldy corn, was also the dietary staple of other poverty-stricken populations in southern Europe that developed pellagra, it was believed that the disease was caused by eating a corn-based diet, particularly when the corn had been improperly stored and became moldy. It was also observed that the diet of those with pellagra was especially lacking in animal protein.

It was not until 1914, when pellagra had become a problem in the southern United States, that Goldberger began his studies, which ultimately demonstrated that the nutritional etiology of pellagra produced a comparable condition in dogs called “black tongue” and identified foods that were pellagra-preventing (Roe 1973). Later investigators extracted some of these foods, including liver, and attempted isolation of nutrients lacking in the diets of those with the disease. During this period, independent studies by Otto Heinrich Warburg and Walter Christian identified nicotinic acid as one of the substances forming “coferment,” which was thought to be important in normal metabolism. Shortly after this discovery, nicotinic acid was found to be an essential factor for the growth of certain bacteria.

Conrad Elvehjem and his co-workers, who carried out extensive studies at the University of Wisconsin in the 1930s, tested various liver-extract fractions for activity as growth factors in experimental animals. One sample was found to contain a substance that was identified by Wayne Woolley as nicotinic acid. Subsequently, others in Elvehjem’s laboratory cured “black tongue” in dogs by feeding them the nicotinic acid obtained from such liver extracts (Kline and Baumann 1971).

Once it was known that nicotinic acid could cure this disease in dogs, Tom Spies gave nicotinic acid to several subjects with pellagra and found that the human disease responded to it as well. The nutrient nicotinic acid was subsequently given the alternate name of niacin in order to avoid lay confusion of the term with nicotine.

Later studies of the metabolism of niacin in experimental animals showed that they could synthesize it

from a precursor, which proved to be the amino acid tryptophan. Because animal proteins are particularly rich in tryptophan, this finding helped to explain the old observation that diets of pellagrins were invariably low in meat and milk (Goldsmith 1964).

Nicotinic acid is active biochemically in several coenzyme forms, and it is necessary for the synthesis and breakdown of fatty acids, carbohydrates, and amino acids. In human studies conducted soon after the discovery of its activity as a vitamin, it was found that when large doses of niacin were administered, individuals became flushed. Although this observation was an early indication that niacin might be toxic if given in excessive amounts, it also suggested that high doses might relieve conditions associated with constriction of blood vessels. In fact, the vitamin failed to yield the hoped-for therapeutic advantage, but it was found later that niacin in high doses could reduce blood cholesterol levels (Havel and Kane 1982).

Riboflavin

Riboflavin, another of the B-complex vitamins, was first isolated from the greenish yellow pigment in milk. Its deficiency occurs in populations whose diets are lacking in dairy foods and green vegetables and produces a disease characterized by cracks at the corners of the mouth, sore tongue, dermatitis of the folds of the body, and changes in the eyes that make them sensitive to light exposure. It was realized that some pellagra sufferers in the United States were deficient in riboflavin as well as in niacin. The condition was also described in very low-income, rural populations in India (Sebrell and Butler 1938). In more recent years, this same condition has been identified in alcoholics. Flavins, including two major active forms of the vitamin designated riboflavin, have also been isolated from liver. Furthermore, it has been shown that flavins are necessary for energy utilization and that requirements for riboflavin increase during periods of physical activity (Roe 1991).

Vitamin B₆

The term “vitamin B₆” embraces a group of closely related compounds: pyridoxine, pyridoxal, and pyridoxamine. Laboratory investigations resulting in the identification of vitamin B₆ were launched after it became known that a particular form of dermatitis, which developed in rats fed purified diets, could not be cured with any known nutrient but could be prevented by giving the animals a particular yeast extract.

The early clinical studies of vitamin B₆ deficiency were based on observations of infants fed an autoclaved formula diet. Then, during the 1950s, when active research was being conducted on the biochemical functions of vitamin B₆, further insight into the clinical manifestations of its deficiency was obtained

by putting human volunteers on a diet that lacked the vitamin and giving them a vitamin B₆ antagonist at the same time.

Following this study, vitamin B₆ deficiency was observed in patients who were given the drug isoniazid for the treatment of tuberculosis. It was then realized that the drug was also a vitamin B₆ antagonist. In both the experimentally produced vitamin B₆ deficiency and the drug-induced disease, dermatitis and neurological symptoms developed, and later an anemia surfaced as well, in which iron was improperly utilized. Biochemical studies showed that there were three physiologically active forms of vitamin B₆ and that these forms are interconvertible. They have been found to function in amino acid metabolism, in the synthesis of heme from which hemoglobin is formed, and in neurotransmitter activity (Roe 1985).

Pantothenic Acid

Metabolic pathways that depend on pantothenic acid include those involved with the biosynthesis of cholesterol and steroid hormones. Abnormalities have been described in several different types of experimental animals placed on diets deficient in pantothenic acid. In rats, these changes included slow rate of growth in young animals and graying of the fur in older animals. The fertility of rats has also been shown to decline, and gait disorders have been observed in pigs. In addition, destruction of the adrenal glands, associated with bleeding into these glands, has been found in postmortem examinations of several species in which the deficiency has been induced.

Little is known of human requirements for the vitamin, but because it is widespread in foods, there is little danger of deficiency. Thus, pantothenic acid deficiency has been described only in humans who are very severely malnourished and in volunteers who have been fed both a pantothenic acid-deficient diet and a pantothenic acid antagonist (Fox 1984).

Biotin

Biotin is another B vitamin that is involved in the biosynthesis of fatty acids. Biotin deficiency has been described in infants as a result of an inborn error of metabolism. A biotin deficiency has also been described in infants and in older individuals with intestinal diseases. Both groups were fed formula diets lacking in biotin, and both were concurrently receiving broad-spectrum antibiotics that prevented the intestinal synthesis of this vitamin (Mock et al. 1985).

Folic Acid (Folacin)

In 1929–31, Lucy Wills and her colleagues described a severe anemia among pregnant women living in conditions of purdah (seclusion) in Bombay. Because of

cultural taboos, these women ate a monotonous and limited diet in which green vegetables and fruits were lacking (Wills and Mehta 1929–30; Wills and Talpade 1930–1). The anemia was characterized by a severe lowering of the red blood cell count and the appearance of large immature cells in the bone marrow and in the peripheral blood. This so-called macrocytic (large-celled) anemia was cured by giving the women a yeast extract. Subsequent studies showed that certain crude liver extracts could also cure it.

Wills also reported poor pregnancy outcomes for the Indian women who were anemic. She subsequently described the anemia in other populations who lived on diets that lacked fruit and green vegetables. Later investigations led to the isolation of a nutrient, both from liver and from green vegetables, that prevented and cured the same macrocytic anemia. This nutrient was named folic acid, or folacin. Early investigations of folic acid showed that it existed in several different forms that had a number of critical functions in cell maturation.

In other studies, researchers deliberately synthesized chemical analogues of folic acid and demonstrated that these substances not only blocked the normal activity of folic acid but also prevented cell division. Such findings led to the development of specific folic acid antagonists, including aminopterin and methotrexate, which inhibit the division of malignant cells. Methotrexate, which was the less toxic of the two substances, is now employed to control certain forms of cancer and leukemia (Roe 1990). However, it was also found that if aminopterin or methotrexate was taken during the first trimester of pregnancy, birth defects occurred in the offspring. The defects were shown in rat studies to develop if one or another of these drugs was administered during the gestational period of embryogenesis (Nelson 1963).

Another association of folic acid nutriture has been revealed more recently, namely that the risk of women bearing infants with neural tube defects, such as spina bifida, is greatly reduced if they are given supplementary folic acid from the time of conception through the first trimester. Currently, a debate continues about whether or not it would be desirable to fortify cereals with folic acid as a public health measure to prevent neural tube defects (MRC Vitamin Study Research Group 1991).

Vitamin B₁₂ (Cobalamin)

Vitamin B₁₂ was first isolated as the active factor in liver extracts that, given by injection, maintained subjects suffering from pernicious anemia. This disease was first described by Thomas Addison in the nineteenth century, who observed that it occurred primarily in older men and women and is characterized by the presence of a macrocytic anemia. The changes in the blood picture are identical to those seen in folic acid deficiency. In addition, patients with pernicious

anemia may develop an irreversible neurological disorder that leads to sensory loss, including loss of balance. A confusional state may also be present.

Studies of pernicious anemia have revealed that it is an autoimmune disease in which cells in the lining of the stomach are destroyed. As a result, the so-called intrinsic factor found in normal gastric secretion is no longer produced in the stomach and there is a loss of capacity to absorb vitamin B₁₂. Vitamin B₁₂ deficiency can also develop when it cannot be absorbed because of surgical loss of the absorption site in the lower part of the small intestine. Aside from these circumstances, a very low intake of vitamin B₁₂ can also cause pernicious anemia, although such a low intake is unlikely to occur except with people who are strict vegans, since the vitamin is present only in animal foods. Deficiency signs may, however, take several years to develop because of body stores of the vitamin built up in earlier periods.

Investigations of the cause of the neurological complications of vitamin B₁₂ deficiency followed recognition that the same complications may develop in those who are exposed to nitrous oxide (Roe 1985). More recently, more such neurological studies have been carried out on the metabolic defect existing in children with genetic disorders of vitamin B₁₂ metabolism (Metz 1993).

Vitamin B₁₂ has essential functions in the synthesis of the amino acid methionine and is also required for the interconversion and cellular uptake of different forms of folic acid (Buchanan 1964).

Daphne A. Roe

Bibliography

- Bender, D. A. 1992. *Nutritional biochemistry of the vitamins*. Cambridge.
- Buchanan, J. M. 1964. The function of vitamin B₁₂ and folic acid co-enzymes in mammalian cells. *Medicine* 43: 697-709.
- Fox, H. M. 1984. Pantothenic acid. In *Handbook of vitamins: Nutritional, biochemical and clinical aspects*, ed. L. J. Machlin, 437-55. New York and Basel.
- Goldsmith, G. A. 1964. The B vitamins: Thiamine, riboflavin, niacin. In *Nutrition: A comprehensive treatise in vitamins, nutrient requirements and food selection*, Vol. 2, ed. G. H. Beaton and E. W. McHenry, 153-4. New York.
- Griminger, P. 1972. Casimir Funk: Biographical sketch. *Journal of Nutrition* 102: 1105-14.
- Havel, R. J., and J. P. Kane. 1982. Therapy in hyperlipidemic states. *Annual Review of Medicine* 33: 417-33.
- Jansen, B. C. P. 1950. Biographical sketch: C. J. Eijkman. *Journal of Nutrition* 42: 1-8.
- Kline, O. L., and C. A. Baumann. 1971. Conrad Arnold Elvehjem: Biographical sketch. *Journal of Nutrition* 101: 569-78.
- Metz, J. 1993. Pathogenesis of cobalamin neuropathy: Deficiency of nervous system S-Adenosylmethionine? *Nutrition Reviews* 51: 12-15.
- Mock, D. M., D. L. Baswell, H. Baker, et al. 1985. Biotin deficiency complicating parenteral alimentation: Diagnosis, metabolic repercussions and treatment. In *Biotin*, ed. K. Dakshinamurti and H. Bgavagan. *Annals of the New York Academy of Sciences* 447: 314-34.
- MRC Vitamin Study Research Group. 1991. Prevention of neural tube defects: Results of the Medical Research Council on vitamin study. *Lancet* 338: 131-7.
- Nelson, M. M. 1963. Teratogenic effects of pteroylglutamic acid deficiency in the rat. In *Ciba Foundation symposium on congenital malformations*, ed. G. E. W. Wolstenholme and C. M. O'Connor, 134-51. Boston, Mass.
- Roe, D. A. 1973. *A plague of corn: The social history of pellagra*. Ithaca, N.Y.
1985. *Drug-induced vitamin deficiencies*. Westport, Conn.
1990. Drug-folate interrelationships: Historical aspects and current concerns. In *Folic acid metabolism in health and disease*, ed. Mary Frances Picciano, E. L. Robert Stokstad, and Jesse F. Gregory III, 277-87. New York.
1991. Ariboflavinosis. In *Hunter's tropical medicine*, Seventh edition, ed. G. T. Strickland and W. B. Saunders, 934-5. Philadelphia, Pa.
- Sebrell, W. H., and R. E. Butler. 1938. Riboflavin deficiency in man: Preliminary note. *Public Health Reports* 53: 2282-4.
- Wills, L., and M. M. Mehta. 1929-30. Studies in pernicious anemia of pregnancy: Part I - Preliminary research. *Indian Journal of Medical Research* 17: 777-8.
- Wills, L., and S. N. Talpade. 1930-1. Studies in pernicious anemia of pregnancy: Part II - A survey of dietetic and hygienic conditions of women in Bombay. *Indian Journal of Medical Research* 18: 307-10.
- Yellowlees, P. M. 1986. Thiamine deficiency and prevention of Wernicke-Korsakoff syndrome. *Medical Journal of Australia* 145: 216-19.

IV.A.3 Vitamin C

In delineating the history of a vitamin, we can often recognize four chronological phases. First, there is the description of a disease of unknown etiology, and second, there is the description of an empirical cure for the disease. Following this step, and often closely associated with it, is the identification of the curative factor - which perforce then becomes known as a vitamin. In the fourth phase the mode of action of the vitamin in preventing the deficiency disease is characterized.

The history of vitamin C (ascorbic acid) conforms to this general pattern. The characterization of the deficiency disease (scurvy) and the empirical discovery of a cure for it are, properly speaking, a part of the history of scurvy and have been dealt with elsewhere in this work. But this chapter is concerned with the subsequent history of the antiscorbutic factor, which conveniently presents itself in three chronological stages: (1) the somewhat ill-defined and open-ended period - from the beginning of the nineteenth century to the 1920s - when the vitamin had the existence of an "unrevealed presence" and was known to exist only because of its preventive influence on the disease scurvy (just as, during the same period, the perturber of Uranus was known to exist long before the "discovery" of the planet Pluto); (2) the 1920s and

the 1930s, when vitamin C was named, isolated, and its molecular structure revealed (in that order); and (3) the modern post-1940 period, with its emphasis on the characterization of the biochemical role of vitamin C in preventing scurvy and, more recently, the debatable "extra-antiscorbutic" roles sometimes attributed to it.

The Antiscorbutic Factor

The early history of vitamin C cannot readily be separated from that of the demise of scurvy. By the beginning of the nineteenth century it was generally accepted that it was possible to prevent and cure scurvy by the use of citrus fruits: James Lind's contribution in establishing this belief was of paramount significance. But it was in essence a pharmacological concept rather than a nutritional one; the belief that the citrus fruits were replacing a missing dietary component would have been alien to medical thought at the beginning of the nineteenth century; even Lind himself did not regard fruit and vegetables as obligatory dietary principles in the prevention of scurvy. In other words, not until the end of the nineteenth century was there any general acceptance that scurvy was a deficiency disease resulting from a lack of a specific dietary principle and that the disease could be prevented or cured by appropriate dietary manipulation. Moreover, even this acceptance was complicated by the advent of the germ theory of disease which, some have argued, caused reversion to an infection theory to explain scurvy's etiology.

One of the earliest thinkers to discuss these new ideas was George Budd (1808–82), Professor of Medicine at King's College, London – although as K. C. Carter has indicated, Budd should perhaps be regarded as a developer rather than as an innovator of the "deficiency disease theory" (Hughes 1973; Carter 1977). In 1842, Budd published in the *London Medical Gazette* a series of articles entitled "Disorders Resulting from Defective Nutriment." He described "three different forms of disease which are already traced to defective nutriment" and argued that such conditions resulted from the absence of dietary factor(s) other than carbohydrate, fat, and protein, and that the absence of each of these specific factors would be associated with a specific disease – an idea that lay in abeyance for some 40 years until experimentally proved by N. Lunin. There can be little doubt that the three diseases described by Budd were avitaminoses A, C, and D.

L. J. Harris, himself a significant figure in the later history of vitamin C, aptly described Budd as "the prophet Budd" and referred to an article in which Budd expressed the belief that scurvy was due to the "lack of an essential element which it is hardly too sanguine to state will be discovered by organic chemistry or the experiments of physiologists in a not too distant future" (Budd 1840; Harris 1937:8).

Little happened, however, to fulfill Budd's prophesy until the beginning of the twentieth century. In 1907, A. Holst and T. Fröhlich of Norway reported experiments in which they had demonstrated that scurvy could be induced in guinea pigs and cured by dietary manipulation (Holst and Fröhlich 1907; Wilson 1975). They used guinea pigs to assess the antiscorbutic value of different foodstuffs and to show the thermolabile nature of the antiscorbutic factor. At the same time there were parallel, but independent, developments in the general theory of vitamin deficiency diseases. F. G. Hopkins, developing earlier work by Lunin, C. A. Pekelharing, W. Stepp, and others, in 1912 published his classic paper in which he demonstrated the presence of growth factors in milk and showed their essential dietary nature (Hopkins 1912); in the same year, Casimir Funk introduced his "vitamin hypothesis," in which he attributed scurvy to the absence of an "anti-scurvy vitamine" (Harris 1937: 1–21).

The use of the guinea pig assay technique for the assessment of the antiscorbutic factor was extended, and in 1917, H. Chick and M. Hume published an important paper in which they reported the factor's distribution in a number of foodstuffs (Chick and Hume 1917). The following year A. Harden and S. S. Zilva published their fractionation studies on lemon juice, in which they demonstrated that the antiscorbutic potency was not attributable (as had been suggested by earlier workers) to the citric acid content (Harden and Zilva 1918). The year after that, J. C. Drummond designated the factor "Water soluble C" (Drummond 1919).

Identification of Vitamin C

Work now began in earnest to identify and isolate the antiscorbutic factor. The Medical Research Council's 1932 publication *Vitamins: A Survey of Present Knowledge* may be referred to for a detailed account of the large number of papers published during the 1920s on what was by then known as "vitamin C." Foremost in these early efforts was Zilva, working at the Lister Institute, London. The essential feature of Zilva's procedure was the precipitation of the factor with basic lead acetate after removal of the bulk of the other organic acids with calcium carbonate. He applied this technique to a variety of sources, such as lemon juice and swede (rutabaga) tissues, and he succeeded in increasing the concentration of the antiscorbutic factor some 200 to 300 times (Hughes 1983).

Other workers were similarly occupied, notably N. Bezssonoff in France and C. G. King in the United States. King has stated that during this period "many investigators had abandoned or failed to publish their work for various reasons." He referred, specifically, to Karl Link of Wisconsin, who had prepared several grams of crude calcium ascorbate during the 1920s but carried his work no further because of lack of financial support (King 1953).

The purest of these early “concentrates” still contained much impurity, and it was not until 1932 that W.A. Waugh and King published their paper “The Isolation and Identification of Vitamin C,” which included a photograph of vitamin C crystals (Waugh and King 1932).

These attempts to isolate and characterize vitamin C were paralleled by two separate, but nevertheless highly relevant, developments in related areas. J. Tillmans and P. Hirsch, German government chemists, extensively studied the capacity of lemon juice preparations to reduce the redox dye 2,6-dichlorophenolindophenol, and they claimed that the reducing power of their preparations was always in proportion to their antiscorbutic potency; the indophenol dye technique later became a standard method for the assay of vitamin C. Zilva, however, disagreed with the German findings and appears, at that point, to have been diverted from his main endeavor by an attempt to disprove them (Zilva 1932).

The other significant development at this time was Albert Szent-Györgyi’s isolation of hexuronic acid. Szent-Györgyi, a Hungarian biochemist working on plant respiration systems at Groningen in Holland, became interested in a reducing compound present in his preparations. Hopkins invited him to Cambridge to extend his studies, and in 1927, Szent-Györgyi isolated his “Groningen reducing agent” in a crystalline, from oranges, lemons, cabbages, and adrenal glands (Szent-Györgyi 1928).

He proposed to name his crystalline sample “ignose” – thus indicating its apparent relationship to sugars while at the same time underlining his ignorance of its true nature. But Harden, the editor of the *Biochemical Journal* at the time, according to Szent-Györgyi, “did not like jokes and reprimanded me.” A second suggestion “godnose” was judged to be equally unacceptable. Szent-Györgyi finally agreed to accept Harden’s somewhat more prosaic suggestion “hexuronic acid” – “since it had 6 Cs and was acidic” (Szent-Györgyi 1963: 1–14).

Hexuronic acid was a strongly reducing compound. So, too, according to Tillmans and Hirsch, was the antiscorbutic substance (vitamin C). The suggestion that hexuronic acid and vitamin C were actually one and the same substance appeared in print in 1932 in papers by both J. L. Svrbely and Szent-Györgyi and by Waugh and King, but there can be little doubt that the idea had been mooted some years previously. Who first made the suggestion is, however, unclear, and even the main participants in the drama later appeared uncertain and confused. King (1953) claimed that it was E. C. Kendall in 1929, but according to Hopkins (reported by King) it was Harris in 1928 (King 1953) – and he had, in any case, already attributed the idea to Tillmans and Hirsch (Harris 1937: 95). But E. L. Hirst (a member of the team later involved in chemical studies on the structure of vitamin C) named Waugh and King (Hirst 1953: 413).

Hopkins had already, in 1928, sent a sample of Szent-Györgyi’s hexuronic acid to Zilva for comments on its vitamin C potency. According to King, Hopkins was disturbed because Zilva (who, naturally perhaps, was reluctant to admit that his “antiscorbutic preparations” were in reality identical with hexuronic acid) had replied that the sample was not vitamin C, but did so without reporting the evidence of his tests (King 1953).

By 1932, however, evidence in favor of the identity of hexuronic acid as vitamin C was substantial. Waugh and King had shown that their “crystalline vitamin C” cured scurvy in guinea pigs (Waugh and King 1932), and earlier the same year, Svrbely and Szent-Györgyi (now working in his native Hungary) had described the antiscorbutic potency of a sample of “hexuronic acid” isolated from adrenal glands (Svrbely and Szent-Györgyi 1932).

In 1933, in a single-sentence letter in *Nature*, Szent-Györgyi and W. N. Haworth drew attention to the chemical inaptness of the term “hexuronic acid” and suggested the term “ascorbic acid,” thus formally acknowledging the antiscorbutic nature of the compound. Harris and his colleagues at Cambridge demonstrated the positive correlation between the hexuronic acid content and the antiscorbutic potency in a wide range of foodstuffs and published a highly convincing “eight-point” proof of the identity of the two substances.

Their three most important points were as follows: (1) Hexuronic acid paralleled antiscorbutic potency; (2) destruction of hexuronic acid by heat or by aeration was accompanied by a corresponding fall in the antiscorbutic activity; and (3) hexuronic acid disappeared from the organs of scorbutic guinea pigs (Birch, Harris, and Ray 1933). There could now be little doubt that hexuronic acid (Tillmans and Hirsch’s reducing compound) and vitamin C were one and the same substance.

The situation was not without its human aspects, and even today the question of priority in the discovery of vitamin C still elicits discussion. “The identification of vitamin C is one of the strangest episodes in the history of vitamins,” wrote T. H. Jukes in commenting on the appearance in 1987 of a book by R. W. Moss that placed, in Jukes’s opinion, too great an emphasis on Szent-Györgyi’s contribution (Moss 1987; Jukes 1988: 1290). Moss had implied that King had rushed off his claim for the identity of vitamin C and hexuronic acid after it became clear to him that Szent-Györgyi intended making the same point in a note to *Nature* – a situation curiously reminiscent of the suggestion that Charles Darwin behaved similarly on learning in 1858 that Alfred Russel Wallace was about to publish his theory of evolution.

The emphasis now shifted to the elucidation of the structure of vitamin C. Haworth, a Birmingham (U.K.) chemist, had received from Szent-Györgyi a sample of his “hexuronic acid,” and in 1933, in a series of impressive papers, the Birmingham chemist, using

both degradative and synthetic procedures, described the structure of the molecule (Hughes 1983). The molecule was synthesized simultaneously, but independently, by T. Reichstein in Switzerland and by Haworth and his colleagues in Birmingham, both groups using essentially the same method.

The synthesis – which, as it later emerged, was quite different from the biosynthetic pathway – was based on the production of xylosone from xylose and its conversion with cyanide to an imino intermediate that, on hydrolysis, gave ascorbic acid. The Swiss group published their results just ahead of the Birmingham workers (Ault et al. 1933, Reichstein, Grüssner, and Oppenheimer 1933). The picture was completed the following year when the Birmingham workers joined forces with Zilva to demonstrate that synthetic ascorbic acid produced at Birmingham had exactly the same antiscorbutic potency as a highly purified “natural” sample from the Lister Institute (Haworth, Hirst, and Zilva 1934).

The annual report of the Chemical Society for 1933, with perhaps unnecessary caution, stated that “although it seems extremely probable that ascorbic acid is vitamin C . . . it cannot be said that this is a certainty.” Other were less circumspect. A. L. Bacharach and E. L. Smith, addressing the Society of Public Analysts and Other Chemists in November 1933, said that “Vitamin C can now be identified with a sugar acid known as ascorbic acid. . . . Contrary to expectation, it is the first vitamin not merely to have assigned to it a definite molecular formula, but actually to be synthesised by purely chemical means” (Hughes 1983). Budd was correct in his 1840 prophecy that both physiologists and chemists could contribute to the identification of the “antiscorbutic factor.” But his “not too distant future” proved to be a period of 93 years!

Biosynthesis and Metabolism of Vitamin C

By the end of the 1930s, serious research had commenced on the biological role of vitamin C. In particular, biochemical reductionists sought to explain the nature of the relationship between the clinical manifestations of scurvy and the biochemical involvements of vitamin C. It was recognized that vitamin C was a powerful biological reductant, and there were early attempts to explain its nutritional significance in terms of its involvement in oxidation-reduction systems – a major theme in prewar biochemistry. But the first clear advance in the biochemistry of vitamin C came from studies of its biosynthesis, and by the early 1950s, the pathway for its formation from simple sugars had been worked out. L. W. Mapson and a colleague at the Low Temperature Research Station at Cambridge (U.K.) fed different possible precursor molecules to cress seedlings and measured the formation of vitamin C. And in the United States, King and co-workers used labeled glucose to chart the biosynthetic pathway in rats (Mapson 1967: 369–80).

The biosynthetic pathway proved to be a comparatively simple one. D-glucuronate (formed from glucose) is converted to L-gluconate and then to L-gulonono-gamma-lactone, which in turn is further reduced (via L-xylo-hexulonolactone) to L-ascorbic acid (2-oxo-L-gulonono-gamm-lactone). The final enzymatic step is catalyzed by L-gulonolactone oxidase (EC 1.1.3.8.) – in the liver in evolutionarily “advanced” species such as the cow, goat, rat, rabbit, and sheep and in the kidney in other species such as the frog, snake, toad, and tortoise – and it is this enzyme that is lacking in those species unable to synthesize vitamin C.

To date, this biochemical “lesion” has been detected in a small, and disparate, number of species – higher primates (including, of course, humans), guinea pigs, certain bats, birds, insects, and fish (Chatterjee 1973; Sato and Uderfriend 1978). Whether all these species are necessarily scurvy-prone is not quite so clear. A survey of 34 species of New World microchiropteran bats showed that L-gulonolactone oxidase was apparently absent from the livers of all of them (and from the kidneys of at least some of them), but nevertheless, the tissue levels of ascorbic acid (even in species that were fish-eaters or insect-eaters) were similar to those in species that could biosynthesize the vitamin (Birney, Jenness, and Ayaz 1976).

This finding would suggest that the vitamin was being synthesized in organs other than the liver and kidney; or that the metabolic requirement for it was remarkably low; or that there were extremely efficient mechanism(s) for its protection against degradative changes. The whole question of the evolutionary significance of vitamin C – in plants as well as in animals – remains a largely uncharted area.

The rate of endogenous biosynthesis of vitamin C in those species capable of producing the vitamin shows considerable interspecies variation, ranging from 40 milligrams (mg) per kilogram (kg) body weight daily for the dog to 275 for the mouse (Levine and Morita 1985). These values are well in excess of the amounts of the vitamin required to prevent the appearance of scurvy in species unable to synthesize it – a finding that has frequently been used to buttress the claim that vitamin C has a number of “extra-antiscorbutic” roles requiring daily intakes well in excess of the recommended daily amounts.

The total body pool of ascorbic acid in a 70 kg man has been estimated at about 1.5 grams (g) (but according to Emil Ginter it could be three times as great as this [Ginter 1980]), which is attainable in most people by the sustained daily intake of 60 to 100 mg. A daily intake of 10 mg vitamin C results in a body pool of about 350 mg. Scorbutic signs do not appear until the pool falls to below 300 mg (Kallner 1981; “Experimental Scurvy” 1986).

Plasma (and less conveniently, leucocyte) concentrations of ascorbic acid are often taken as an index of the body status of the vitamin. The normal concentra-

tion range in the plasma of healthy persons on an adequate plane of nutrition is 30 to 90 micromoles per liter ($\mu\text{mol/L}$) (0.5–1.6 mg/100 ml). The Nutrition Canada Interpretive Guidelines are often referred to in this respect; these guidelines suggest that values between 11 and 23 $\mu\text{mol/L}$ are indicative of marginal deficiency and that values below 11 $\mu\text{mol/L}$ point to frank severe deficiency – but differences of sex, race, metabolism, smoking habits, and, particularly, of age (factors known to influence plasma ascorbic acid concentrations) reduce the validity of such a generalization (Basu and Schorah 1981; Hughes 1981b).

During a period of vitamin C depletion there is a comparatively rapid loss of vitamin C (a reduction of about 3 percent in the body pool daily) resulting from the continued catabolism of the vitamin and the excretion of its breakdown products in the urine. In humans, the main pathway identified involves the conversion of the ascorbic acid to dehydroascorbic acid, diketogulonic acid, and oxalic acid (in that order), with the two latter compounds accounting for the bulk of the urinary excretion of breakdown products. Smaller amounts of other metabolites, such as ascorbic acid-2-sulphate also occur, and in the guinea pig there is substantial conversion of part of the ascorbic acid to respiratory CO_2 . It has sometimes been argued that the excess formation of these catabolites (particularly oxalic acid) should signify caution in the intake of amounts of vitamin C substantially in excess of the amount required to prevent scurvy.

Biochemical Role of Vitamin C

It was noted in the early experiments of Holst and Fröhlich, and confirmed by many subsequent workers, that defective formation of connective tissue was a primary pathological feature of experimental scurvy, and at one time it was believed that this lesion could account for most of the known pathological sequelae of the disease – the petechial hemorrhages, the breakdown of gum tissue, and the impairment of wound repair tissue. Attempts to characterize the biochemical *modus operandi* of vitamin C in preventing scurvy, therefore, centered initially on the metabolism of collagen – the essential glycoprotein component responsible for imparting strength to connective tissue.

By the 1970s, there was suggestive evidence that the biochemical lesion was located in the hydroxylation of the proline and lysine components of the collagen polypeptide and that vitamin C had an essential role in the process (Barnes and Kodicek 1972). The hydroxylases involved in collagen biosynthesis (prolyl 4-hydroxylase, prolyl 3-hydroxylase, and lysyl hydroxylase) require ferrous iron as a cofactor, and it appears that vitamin C, a powerful biological reductant, has an almost obligatory role in maintaining the ferrous iron in the reduced form. Thus emerged a simplistic and reductionist explanation for the role of vitamin C in

preventing the emergence of the main clinical features of scurvy.

Yet although there can be little doubt that vitamin C plays a critical role in the biosynthesis of collagen, recent studies have suggested that the simple “defective hydroxylation” theory is, perhaps, not the complete story. Studies have indicated that the activity of prolyl hydroxylase and the formation of collagens by fibroblast cultures is not influenced by ascorbic acid; furthermore, ascorbic acid deficiency does not always result in severe underhydroxylation of collagen in scorbutic guinea pigs (England and Seifter 1986).

There is increasing evidence that vitamin C may also influence the formation of connective tissue by modifying the nature and formation of the extracellular matrix molecules (Vitamin C Regulation 1990). B. Peterkofsky (1991) has recently suggested that the role of vitamin C in collagen biosynthesis is a dual one – a direct influence on collagen synthesis and an indirect one (mediated perhaps via appetite) on proteoglycan formation.

The complement component C1q, which has a central role in disease resistance, contains a collagen-like segment that is rich in hydroxyproline, and it has been suggested that this segment could offer a link with the putative anti-infective powers widely suggested for vitamin C (Pauling 1976; see also the section on megatherapy). Studies over the last 15 years, however, have failed to demonstrate that the complement system, unlike connective tissue collagen, reflects vitamin C availability (Thomas and Holt 1978; Johnston 1991). Indeed, the belief that vitamin C had anti-infection powers probably stemmed from reports by Harris in 1937 of lowered vitamin C in persons suffering from certain diseases, particularly tuberculosis.

During the 1960s and the 1970s, however, some 25 epidemiological studies were completed in different parts of the world to assess the validity of claims that vitamin C had anti-infection powers, particularly with respect to the common cold. The general conclusion drawn from the results of these studies was that the evidence for a protective/curative role for vitamin C in the common cold was far from convincing (Hughes 1981b: 22–6; Carpenter 1986: 213–16).

There is accumulating evidence, however, that vitamin C may have additional involvements in a range of enzymatic changes unrelated to the formation of collagen. There are three systems of considerable physiological significance in which vitamin C plays an important, and possibly obligatory, role: (1) as the immediate donor for dopamine B-hydroxylase, a key reaction in the conversion of tyrosine to norepinephrine (England and Seifter 1986; Fleming and Kent 1991); (2) in the peptidylglycine α -amidating monooxygenase system, whereby peptidyl carboxyl-terminal residues are amidated, a process that requires molecular oxygen, copper, and ascorbate and is important in the biosynthesis of a number of neuroendocrine peptides (England and Seifter 1986; Eipper and Mains 1991); (3)

in the hydroxylation reactions in the biosynthesis of carnitine from lysine and methionine (England and Seifter 1986; Rebouche 1991).

The exact physiological significance of these and other reactions vis-à-vis the clinical manifestations of scurvy is unclear. The first two, having obvious involvements in the endocrine and nervous systems, could well be causally related to various functional derangements of scurvy; and as carnitine has an important role in the transport of fatty acids into the mitochondria, where they may be oxidized to provide energy, it has been suggested that the carnitine involvement could account for the lassitude and fatigue that have been invariably noted as an early feature of scurvy (Hughes 1981a).

Should such involvements require an availability of ascorbic acid greater than that required to prevent the emergence of "classical" scurvy – and there is some evidence that this is so, at least in the case of carnitine biosynthesis – then a revision of the currently accepted Recommended Dietary Allowance/Reference Value would be called for (Hughes 1981a). The current recommended daily intake of vitamin C (60 mg in the United States and recently raised from 30 to 40 in the United Kingdom) is, after all, the amount estimated to prevent the emergence of the classic ("collagen") features of scurvy – and in the United Kingdom, it is based, essentially, on a single experiment completed almost half a century ago on a non-representative population sample.

Source of Vitamin C

Vitamin C is a heat-labile, water-soluble, and readily oxidizable molecule, and its distribution among foodstuffs and the losses resulting from processing and food preparation have been well documented. Studying the losses induced in the vitamin C content of various foodstuffs by simple culinary procedures must be one of the commonest and oft-repeated projects in basic college and university courses, and the amount of unpublished data resulting from these studies must be immense.

The mean daily intake of vitamin C in the United Kingdom (based on noncooked purchases) is about 60 mg daily with potatoes, citrus fruits, and cabbage accounting for 20, 12, and 6 percent, respectively, of the intake. The losses resulting from cooking are substantial, and these are further increased if the cooked food is allowed to stand around before being eaten. Nevertheless, because of the comparatively widespread distribution of the vitamin in plant foodstuffs, and the role of technology in increasing the availability of uncooked plant and vegetable material during the whole year, very few persons today appear to suffer from clinically defined hypovitaminosis C; consequently, frank scurvy is an almost unknown condition.

A recent survey of vitamin C intakes in European countries revealed an interesting, and almost provi-

dential, reciprocity between the consumption of two important sources of vitamin C. Of 27 countries studied, Iceland, Switzerland, and France had the lowest annual consumption of cabbage (less than 5 kg per capita) but a high consumption of citrus fruit (over 20 kg); Romania, Poland, and the former Soviet Union, in contrast, had the lowest consumption of citrus fruit (less than 4 kg) but the highest consumption of cabbage (more than 30 kg) (Kohlmeier and Dortschy 1991). Only where a person, for ideological, economic, or supposed "health" reasons subsists on a diet devoid of fruit and vegetables (such as one based on nuts, grain, and/or cooked meat/fish) is scurvy likely to emerge.

The foliage of many flowering plants has an unexpectedly high concentration of vitamin C, with concentrations of up to 1 percent wet weight being attained in some members of the Primulaceae family. The mean concentration for 213 species examined (162 mg per 100 g) was some three times that of those culinary vegetables usually regarded as good sources of the vitamin; and the mean value for the leaves of 41 woody shrubs and trees examined was 293 mg per 100 g – significantly higher than black currants, which are usually cited as the dietary source par excellence of vitamin C (Jones and Hughes 1983, 1984; see also Table IV.A.3.1).

Table IV.A.3.1. *Ascorbic acid content of some plants (mg/100 g fresh weight)*

Culinary vegetables and fruits	
Cabbage	55 (11)
Carrots (roots)	8 (2)
Cauliflower	70 (14)
Peas (seed)	25 (4)
Potatoes (tubers)	28 (4)
Apples (fruit)	10
Black currants (fruit)	180
Oranges (fruit)	55
The antiscorbutics (leaves)	
<i>Cochlearia officinalis</i> (scurvy grass)	63
<i>Nasturtium officinale</i> (watercress)	83
<i>Veronica beccabunga</i> (brooklime)	46
Less widely used antiscorbutics	
<i>Galium aparine</i> (goosegrass)	84
<i>Menyanthes trifoliata</i> (bog bean)	74
<i>Urtica dioica</i> (stinging nettle)	169
Vitamin C-rich leaves	
<i>Primula vulgaris</i> (primrose)	805
<i>Malus domestica</i> (apple tree)	496
<i>Erica tetralix</i> (bog heather)	394
<i>Vicia cracca</i> (tufted vetch)	349

Note: Values in parentheses are for cooked meals taken by elderly patients in hospitals (Jones, Hughes, and Davies 1988 and additional unpublished material).

Source: Based on Jones and Hughes (1983, 1984) and Hughes (1990).

The historically important “antiscorbutic herbs” are among the poorest sources of vitamin C (Hughes 1990). William Perry, who stewed boxes of mustard greens and cress on board his ship in an attempt to fend off scurvy during his Arctic expedition of 1818 (Lloyd and Coulter 1963: 108), would have done better to adorn his stateroom with primrose plants, a single leaf of which, chewed in the mouth daily, would have sufficed to offer complete protection. The exact reason, if any, for these high (and often disparate) concentrations of ascorbic acid in angiosperms is not known; nor is the role of ascorbic acid in plant biochemistry understood. It has been suggested that there is a positive correlation between the concentration of ascorbic acid in plants and corresponding concentrations of phenolic compounds, but the extent to which this reflects a biochemical relationship is a matter of conjecture (E. C. Bates-Smith, personal communication).

Vitamin C Megatherapy

The practice of ingesting daily doses of vitamin C grossly in excess of the amount believed to protect against scurvy and even in excess of the amount known to produce tissue saturation is one of the more controversial aspects of current nutritional thought. The arguments for vitamin C “megatherapy” were initially outlined in the United States by Irwin Stone and later elaborated by Linus Pauling, winner of two Nobel Prizes (Hughes 1981b: 47–53). Stone disputed the adequacy of current recommended daily intakes, basing his case primarily on the rate of biosynthesis of the vitamin by animals producing their own supply and on G. H. Bourne’s estimate that the natural diet of the gorilla provides it with a daily intake of some 4.5 g ascorbic acid (Bourne 1949). His arguments for daily intakes of grams rather than milligrams were enthusiastically embraced and extended by Pauling.

Closely interwoven with the megatherapy theory is the claim that vitamin C has a number of extra-antiscorbutic functions (protection against infection and, particularly, the common cold, detoxication, cerebral function, lipid metabolism, longevity, and so forth) that might require significantly raised amounts of the vitamin (Hughes 1981b: 14–34). For example, E. Ginter has for many years carefully presented the thesis that vitamin C plays a part in lipid metabolism, particularly by enhancing the conversion of cholesterol to bile salts, and that it would, therefore, have a hypocholesterogenic function (Ginter and Bobek 1981).

To date, however, there is little evidence that these putative relationships are reflected by a specific and increased demand for vitamin C. And as indicated earlier, some of these supposed secondary roles have now been subsumed in enzymatic terms by the advances of reductionist biochemistry. Secondary (or extra-antiscorbutic) roles for vitamin C could, conceivably,

require intakes greater than those necessary for the prevention of classical scurvy, but such increased requirements would, in biochemical terms, scarcely justify the massive intakes recommended by the megatherapists.

Apart from the lack of satisfactory evidence, there are other arguments against vitamin C megatherapy (Jukes 1974; Hughes 1981b: 47–53). Adverse reactions elicited by massive doses of vitamin C and the possibly toxic influence of its breakdown products could well disadvantage the body. Moreover, the ingestion of large amounts of ascorbic acid is a self-defeating exercise as the absorption of large doses is a relatively inefficient process, with less than one-half of a 1 g megadose being absorbed from the gastrointestinal tract and only one-fourth of a 5 g dose (Davies et al. 1984; “Experimental Scurvy” 1986). And, in any case, it is generally accepted that tissue saturation in humans may be satisfactorily attained by a daily intake of 100 to 150 mg or even less. The faith of the megatherapists would have appeared to blind them to the normal canons of scientific assessment.

In the mid-1970s, Pauling espoused perhaps the most controversial of all his vitamin C beliefs. In collaboration with a Scottish surgeon, Ewan Cameron, he began to write extensively on the supposed antitumor activity of vitamin C; more specifically, Cameron and Pauling published the results of a clinical trial in which it was claimed that a megadose (10 g daily) of vitamin C quadrupled the survival time of terminally ill cancer patients (Cameron and Pauling 1976). The methodology of this trial was widely criticized, and a carefully controlled attempt to repeat it at the Mayo Clinic in the United States failed to confirm the Cameron–Pauling claims. For the next 15 years, and in the face of growing reluctance on the part of the scientific press to publish his papers, Pauling continued to present his arguments for the efficacy of vitamin C in the treatment of cancer. An account of this drawn-out battle between Pauling and the American scientific establishment has recently appeared (Richards 1991).

In more general and theoretical terms, it has been suggested that the antioxidant and free-radical scavenger roles of vitamin C support its possible function in the prevention (as contrasted with the cure) of cancer. G. Block has assessed some 90 studies of cancer and vitamin C/fruit intake relationships and has concluded that there is evidence that in the majority of cancers vitamin C may have a significant prophylactic role (Block 1991). In this respect, the possible relationship between vitamin C and nitrosamine-induced cancers has attracted some attention.

It has been speculated that endogenously produced N-nitroso compounds may be important initiators of human cancers. Significant in this respect is the formation of N-nitrosamines and related compounds. Nitrosamines may be formed when nitrate, a suitable “nitrosable” amine, and bacteria coexist – as

in the gastrointestinal tract. Nitrate (the main dietary sources of which are fish and root vegetables) is converted by bacterial action to nitrite, which then reacts with amines to produce carcinogenic nitrosamines. Some foods, particularly cured meat products, contain nitrosamines formed during processing.

There is evidence that vitamin C may prevent the formation of carcinogenic nitrosamines from nitrate and may even reduce the carcinogenicity of preformed nitrosamines (Hughes 1981b: 27-9; MAFF 1987). It has been suggested, for example, that a reduction in nitrosamine formation, attributable to citrus fruit vitamin C, may be a contributory factor in determining the comparatively low incidence of large bowel cancer in the "citrus belt" of the United States (Lyko and Hartmann 1980). Sodium nitrite is used in the large-scale preparation of cured meats, bacon, and sausages (primarily to prevent the activity of the highly toxic *Clostridium botulinum*), and these products, consequently, contain a range of preformed nitroso compounds. There may, therefore, be good scientific reasons for regarding orange juice as a useful dietary accompaniment to a fried breakfast!

Vitamin C and Industry

Many tens of thousands of tons of vitamin C are produced synthetically each year from glucose; the initial stages in conversion involve reduction to sorbitol followed by bacterial oxidation to sorbose. Much of this vitamin C finds its way into health-food stores for sale to megatherapy enthusiasts as a putative dietary adjuvant. A substantial proportion is used industrially as a "technological aid"; some is used in meat-curing processes to promote pigment conversion (in this application, it also has an adventitious and unintended role in reducing the formation of volatile N-nitrosamines).

Vitamin C is also widely employed as a permitted antioxidant (sometimes as ascorbyl palmitate) to prevent the formation of rancidity in stored fat products and the phenolic browning of commodities such as dehydrated potatoes. It is used as a flour improver in the Chorleywood Bread Process, where its oxidation product (dehydroascorbic acid) modifies the availability of glutathione in dough development, thereby shortening the period of fermentation.

The use of vitamin C in these technological processes finds general approval on the grounds that one is, after all, adding a beneficial vitamin rather than some untried additive of unknown toxicity. It should be pointed out, though, that little of this additive vitamin C is recoverable from the marketed product, which will, however, contain substantial amounts of vitamin C breakdown products - many of them unidentified and almost all of them of unknown toxicity. Bread is vitamin C-free despite the substantial amounts that may have been added during the Chorleywood process. It has been estimated that the aver-

age consumer may ingest up to 200 mg a week of vitamin C breakdown products from additive sources (Thomas and Hughes 1985).

Epilogue

Why vitamin C should have attracted so much attention in nutritional circles - orthodox and otherwise - is difficult to understand. Its almost limitless appeal to health enthusiasts and pseudonutritionists is matched only by the time and attention devoted to it by academic nutritionists. The annual global publication of some 2,000 papers bearing on vitamin C implies an annual research expenditure of some £40,000,000 (about 60 to 70 million U.S. dollars) - a not inconsiderable sum for studying a molecule whose nutritional significance is, at the most, marginal. Vitamin C deficiency is today a rare occurrence, and the evidence for extra-antiscorbutic requirements in excess of the mean daily intake is slender. Perhaps the biochemical versatility of the vitamin C molecule makes it attractive to biochemists who feel that it deserves a much more significant role than that of a somewhat prosaic involvement in the biosynthesis of collagen.

There are some questions that remain unanswered. For example, the apparent negative correlation between blood and tissue concentrations of vitamin C and age is puzzling. Many very elderly subjects - particularly if institutionalized - have virtually no ascorbic acid in their blood, a situation that would almost certainly be associated with the emergence of clinical scurvy in a younger age group. Yet these octogenarians and nonagenarians seem to be in no way disadvantaged by the apparent absence of the vitamin. Is there, then, a negative correlation between aging and dependency upon vitamin C? Such a relationship, if true, would be a remarkably fortuitous one as a substantial proportion of the institutionalized elderly have intakes of vitamin C well below the recommended daily amount. It is a somewhat sobering thought that in these days of scientifically attuned dietetics the mean intake of vitamin C by the institutionalized elderly in the United Kingdom is no greater than it was in hospitals a century and a half ago (Jones, Hughes, and Davies 1988).

In the more rarefied atmosphere of academic biochemistry, however, it is possible to point to real advances in our knowledge of vitamin C over the past 40 years. Today, modern high-performance liquid chromatographic techniques are replacing the classical indophenol dye method for the determination of vitamin C, with an increase in sensitivity and specificity. Our knowledge of possible biochemical involvements of the vitamin has advanced substantially. Sadly, however, one cannot point with equal certainty to any corresponding expansion of our knowledge of the nutritional significance of vitamin C beyond its role in the prevention of classical scurvy.

R. E. Hughes

Bibliography

- Ault, R. G., D. K. Baird, H. C. Carrington, et al. 1933. Synthesis of d- and L-ascorbic acid and of analogous substances. *Journal of the Chemical Society* 1: 1419-23.
- Barnes, M. J., and E. Kodicek. 1972. Biological hydroxylations and ascorbic acid with special regard to collagen metabolism. *Vitamins and Hormones* 30: 1-43.
- Basu, T. K., and C. J. Schorah. 1981. *Vitamin C in health and disease*. London.
- Birch, T. W., L. J. Harris, and S. N. Ray. 1933. Hexuronic (ascorbic) acid as the antiscorbutic factor, and its chemical determination. *Nature* 131: 273-4.
- Birney, E. C., R. Jenness, and K. M. Ayaz. 1976. Inability of bats to synthesize L-ascorbic acid. *Nature* 260: 626-8.
- Block, G. 1991. Epidemiologic evidence regarding vitamin C and cancer. *American Journal of Clinical Nutrition* 54: 1310S-14S.
- Bourne, G. H. 1949. Vitamin C and immunity. *British Journal of Nutrition* 2: 346-56.
- Budd, G. 1840. Scurvy. In *The library of medicine. Practical medicine*, Vol. 5, ed. A. Tweedie, 58-95. London.
1842. Disorders resulting from defective nutriment. *London Medical Gazette* 2: 632-6, 712-16, 743-9, 906-15.
- Cameron, E., and L. Pauling. 1976. Supplemental ascorbic acid in the supportive treatment of cancer: Prolongation of survival time in terminal human cancer. *Proceedings of the National Academy of Science* 73: 3685-9.
- Carpenter, K. J. 1986. *The history of scurvy and vitamin C*. Cambridge and New York.
- Carter, K. C. 1977. The germ theory, beriberi, and the deficiency theory of disease. *Medical History* 21: 119-36.
- Chatterjee, I. B. 1973. Evolution and biosynthesis of ascorbic acid. *Science* 182: 1271-2.
- Chick, H., and M. Hume. 1917. The distribution among food-stuffs . . . of the substances required for the prevention of (a) beriberi and (b) scurvy. *Transactions of the Royal Society for Tropical Medicine and Hygiene* 10: 141-86.
- Davies, H. E. F., J. E. W. Davies, R. E. Hughes, and Eleri Jones. 1984. Studies on the absorption of L-xylo ascorbic acid (vitamin C) in young and elderly subjects. *Human Nutrition: Clinical Nutrition* 38: 469-71.
- Drummond, J. C. 1919. Note on the role of the anti-scorbutic factor in nutrition. *Biochemical Journal* 13: 77-88.
- Eipper, B. A., and R. E. Mains. 1991. The role of ascorbate in the biosynthesis of neuroendocrine peptides. *American Journal of Clinical Nutrition* 54: 1153S-6S.
- Englard, S., and S. Seifter. 1986. The biochemical functions of ascorbic acid. *Annual Review of Nutrition* 6: 365-406.
- Experimental scurvy in a young man. 1986. *Nutrition Reviews* 44: 13-5.
- Fleming, P. J., and U. M. Kent. 1991. Cytochrome b561, ascorbic acid, and transmembrane electron transfer. *American Journal of Clinical Nutrition* 54: 1173S-8S.
- Ginter, E. 1980. What is truly the maximum body pool of ascorbic acid in man? *American Journal of Clinical Nutrition* 33: 538-9.
- Ginter, E., and P. Bobek. 1981. The influence of vitamin C on lipid metabolism. In *Vitamin C (ascorbic acid)*, ed. J. N. Counsell and D. H. Hornig. London.
- Harden, A., and S. S. Zilva. 1918. The antiscorbutic factor in lemon juice. *Biochemical Journal* 12: 259-69.
- Harris, L. J. 1937. *Vitamins in theory and practice*. New York and Cambridge.
1953. Chairman's summing-up. *Proceedings of the Nutrition Society* 12: 341-4.
- Haworth, W. N., E. L. Hirst, and S. S. Zilva. 1934. Physiological activity of synthetic ascorbic acid. *Journal of the Chemical Society* 2: 1155-6.
- Hirst, E. L. 1953. The chemistry of vitamin C. In *Lind's treatise on scurvy*, ed. C. P. Stewart and D. Guthrie, 413-24. Edinburgh.
- Holst, A., and T. Fröhlich. 1907. Experimental studies relating to ship beri-beri and scurvy. II. On the etiology of scurvy. *Journal of Hygiene* 7: 634-71.
- Hopkins, F. G. 1912. Feeding experiments illustrating the importance of accessory factors in normal diets. *Journal of Physiology* 44: 425-60.
- Hughes, R. E. 1973. George Budd (1808-1882) and nutritional deficiency diseases. *Medical History* 17: 127-35.
- 1981a. Recommended daily amounts and biochemical roles - the vitamin C, carnitine, fatigue relationship. In *Vitamin C (ascorbic acid)*, ed. J. N. Counsell and D. H. Hornig. London.
- 1981b. *Vitamin C: Some current problems*. London.
1983. From ignose to hexuronic acid to vitamin C. *Trends in Biochemical Science* 8: 146-7.
1990. The rise and fall of the "antiscorbutics": Some notes on the traditional cures for "land scurvy." *Medical History* 34: 52-64.
- Johnston, C. S. 1991. Complement component C1q unaltered by ascorbate supplementation in healthy men and women. *Nutritional Biochemistry* 2: 499-501.
- Jones, E., and R. E. Hughes. 1983. Foliar ascorbic acid in some angiosperms. *Phytochemistry* 22: 2493-9.
1984. A note on the ascorbic acid content of some trees and woody shrubs. *Phytochemistry* 23: 2366-7.
- Jones, E., R. E. Hughes, and H. E. F. Davies. 1988. Intake of vitamin C and other nutrients by elderly patients receiving a hospital diet. *Journal of Human Nutrition and Dietetics* 1: 347-53.
- Jukes, T. H. 1974. Are recommended and daily allowances for vitamin C adequate? *Proceedings of the National Academy of Science* 71: 1949-51.
1988. The identification of vitamin C, an historical summary. *Journal of Nutrition* 118: 1290-3.
- Kallner, A. 1981. Vitamin C - man's requirements. In *Vitamin C (ascorbic acid)*, ed. J. N. Counsell and D. H. Hornig. London.
- King, C. G. 1953. The discovery and chemistry of vitamin C. *Proceedings of the Nutrition Society* 12: 219-27.
- Kohlmeier, L., and R. Dortschy. 1991. Diet and disease in East and West Germany. *Proceedings of the Nutrition Society* 50: 719-27.
- Levine, M., and K. Morita. 1985. Ascorbic acid in endocrine systems. *Vitamins and Hormones* 42: 1-64.
- Lloyd, H. C., and J. L. S. Coulter. 1963. *Medicine and the navy, 1200-1900*, Vol. 4. Edinburgh and London.
- Lyko, H. C., and J. X. Hartmann. 1980. Ascorbate, cyclic nucleotides, citrus and a model for preventing large bowel cancer. *Journal of Theoretical Biology* 83: 675-86.
- Mapson, L. W. 1967. Biogenesis of L-ascorbic acid in plants and animals. In *The Vitamins*, Vol. 1, ed. W. H. Sebrell and R. S. Harris, 369-83. New York and London.
- Ministry of Agriculture, Fisheries, and Food (MAFF). 1987. *Nitrate, nitrite and N-nitroso compounds in food*. London.
- Moss, R. W. 1987. *Free radical: Albert Szent-Györgyi and the battle over vitamin C*. New York.
- Pauling, L. 1976. *Vitamin C, the common cold and the flu*. San Francisco.
- Peterkofsky, B. 1991. Ascorbate requirement for hydroxyla-

- tion and secretion of procollagen: Relationship to inhibition of collagen synthesis in scurvy. *American Journal of Clinical Nutrition* 54: 1113S-27S.
- Rebouche, C. J. 1991. Ascorbic acid and carnitine biosynthesis. *American Journal of Clinical Nutrition* 54: 1147S-52S.
- Reichstein, T., A. Grussner, and R. Oppenheimer. 1933. Synthesis of d- and L-ascorbic acid (vitamin C). *Nature* 132: 280.
- Richards, E. 1991. *Vitamin C and cancer*. London.
- Sato, P., and S. Uderfriend. 1978. Studies on ascorbic acid related to the genetic basis of scurvy. *Vitamins and Hormones* 36: 33-52.
- Svirbely, J. L., and A. Szent-Györgyi. 1932. Hexuronic acid as the antiscorbutic factor. *Nature* 129: 576.
- Szent-Györgyi, A. 1928. Observations on the function of the peroxidase systems. . . . *Biochemical Journal* 22: 1387-1409.
1963. Lost in the twentieth century. *Annual Review of Biochemistry* 32: 1-14.
- Thomas, M., and R. E. Hughes. 1985. Evaluation of threonic acid toxicity in small animals. *Food Chemistry* 17: 79-83.
- Thomas, W. R., and P. G. Holt. 1978. Vitamin C and immunity: An assessment of the evidence. *Clinical and Experimental Immunology* 32: 370-9.
- Vitamin C regulation of cartilage modification. *Nutritional Reviews* 48: 260-2.
- Waugh, W. A., and C. G. King. 1932. The isolation and identification of vitamin C. *Journal of Biological Chemistry* 97: 325-31.
- Wilson, L. G. 1975. The clinical definition of scurvy and the discovery of vitamin C. *Journal of the History of Medicine* 30: 40-60.
- Zilva, S. S. 1932. Hexuronic acid as the antiscorbutic factor. *Nature* 129: 690.

IV.A.4 Vitamin D

Definition and Nomenclature

Vitamin D is a fat-soluble substance required by most vertebrates, including humans, to keep blood calcium and phosphate levels within a narrow normal range and thereby maintain a normal skeleton and optimal cellular function. The term, vitamin D, is a misnomer. Vitamin D is not a vitamin. It is synthesized in the skin, and so, unlike other vitamins, which are essential dietary components, it does not satisfy the criteria for classification as a vitamin. Nor is it a hormone because it is biologically inactive and must be metabolized by the body into a multihydroxylated version, known as calcitriol, which is biologically active and the true hormonal form. Thus vitamin D is more accurately described as a *prohormone*. The natural form of the vitamin, known as vitamin D₃, is a cholesterol-like substance produced in the skin by a nonenzymatic process involving ultraviolet light and heat. An artificial form of the vitamin, with an altered side chain, known as vitamin D₂, is derived from the plant

sterol ergosterol and is often used instead of vitamin D₃ as a dietary supplement.

Most of the complexity associated with the nomenclature in the vitamin D field stems from confusion surrounding its discovery during the period 1919 to 1922. Early research showed that the deficiency associated with lack of vitamin D (rickets in children or osteomalacia in adults) was cured by seemingly unrelated treatments: exposure to sunlight or ingestion of a fat-soluble substance. The early nutritional pioneers of that period, including Sir Edward Mellanby and Elmer V. McCollum, realized that several related factors would cure rickets and that one of these substances, vitamin D₃, could be made in the skin. Students often ponder the fate of vitamin D₁. It was a short-lived research entity comprising a mixture of vitamins D₂ and D₃, and the term has no value today. Vitamin D₃ is sometimes referred to as cholecalciferol or, more recently, calcitriol; vitamin D₂ is known as ergocalciferol or ercalciol. The discovery of the hydroxylated versions of vitamin D by Hector E. DeLuca and Egon Kodicek in the 1967 to 1971 period led to a major expansion of our knowledge of a number of biologically active compounds, but calcitriol is the singularly most important version of these. For purposes of discussing the history of foodstuffs, we shall use the term vitamin D to describe all substances that can be activated to produce biological effects on calcium and phosphate metabolism in humans.

History of Vitamin D Deficiency (Rickets)

Though the nutritional entity vitamin D has been known for only 75 years, the deficiency diseases of vitamin D (rickets and its adult-onset counterpart, osteomalacia) were clearly recognized by Daniel Whistler (1645) in the Netherlands and Francis Glisson (1650) in England as early as the mid-seventeenth century.

In reviewing the history of rickets, we must recognize that rickets and osteomalacia are not the only diseases that affect the skeleton. Others include osteoporosis, which results from loss of total bone (i.e., proteinaceous matrix and minerals), and hormonal imbalances (e.g., hyperparathyroidism). Because diagnostic procedures were primitive until the twentieth century, the term "rickets" may often have been applied to other skeletal abnormalities and conditions not caused by vitamin D deficiency. Nevertheless, in many cases there is sufficient detail in the descriptions provided to recognize the condition.

According to strict medical classification, rickets and osteomalacia encompass a group of skeletal malformations resulting from a spectrum of different causes but having the common feature that *the bone matrix is insufficiently mineralized or calcified*. By far the most common cause of rickets is a lack of vitamin D. This deficiency must be the result of inad-

quate skin synthesis of vitamin D₃ compounded by low dietary intake of vitamin D. The term “rickets” is thought by most to have its origins in the verb in the Dorset dialect *to rucket*, which means to breathe with difficulty. Yet some claim that the term is derived from the Anglo-Saxon word *wrikken*, meaning to twist.

Rickets is characterized by a deformed and misshaped skeleton, particularly bending or bowing of the long bones and enlargement of the epiphyses of the joints of the rib cage, arms, legs, and neck. Victims have painful movements of the rib cage and hence difficulty breathing. In China, medical texts refer to deformities of the rib cage in severe rickets as “chicken breast.” Severe rickets is often accompanied by pneumonia. There is currently much research in progress on a second important function of vitamin D, namely to control the differentiation and development of cells of the bone marrow and immune system. Thus with rickets, the defects of the skeleton may be accompanied by reduced ability to fight infections. Rachitic patients have difficulty holding up their heads, which is sometimes depicted in lithographs from the period circa 1650 to 1700 (e.g., Glisson’s “De Rachitide,” 1650). A more thorough review of the history of rickets can be found in the extraordinarily detailed book of Alfred Hess (1929) entitled *Rickets Including Osteomalacia and Tetany*. Though rickets is rarely life-threatening, it certainly lowers the quality of life for the afflicted individual and probably leads to secondary problems. One of the best documented of these secondary problems is the development of deformities of the pelvis in young females, which can cause difficulties in childbirth. This topic has been given detailed analysis by the University of Toronto historian Edward Shorter (1982) in *A History of Women’s Bodies*. Shorter concludes that before 1920, women who had contracted rickets earlier in life had the risk of a “contracted pelvis” that must have caused numerous deaths during their first delivery.

Discovery of Vitamin D

Around the turn of the twentieth century, several physicians noted a seasonal variation in the incidence of rickets and that the disease was associated with lack of exposure to sunlight. In fact, several researchers also noted a higher incidence of rickets in the industrialized cities of northern Europe (which lay under a pall of smoke caused by burning coal in open fires and factories) than in the rural areas around these centers. The Dickensian character Tiny Tim, of the novel *A Christmas Carol*, clearly represents a child who must have been a common sight in the narrow alleyways of the dark cities of the late nineteenth century. In retrospect, it is easy to see how rickets could be prevalent among the occupants of the sweatshops of such dingy cities when their agrarian cousins on a similar diet, but out in the sun 12 hours a day, had no rickets. Sunbaths were recom-

mended by Edwardian physicians as a cure for this condition, but some believed that the accompanying “fresh-air and exercise,” rather than sunlight, were the key ingredients in the cure. In 1912, J. Raczynski (1913; described in Hess 1929) performed a definitive experiment by exposing two rachitic puppies to either sunlight or shade for six weeks and showing that the sunlight-exposed animal had a 1.5-fold higher bone mineral content. Nevertheless, controversy persisted as other researchers claimed to show the importance of “country-air and exercise” in the prevention of rickets and still others showed that the onset of rickets could be accelerated by dietary manipulation. Yet it is now clear that a “rachitogenic diet” (a diet that can lead to rickets) is one that contains adequate amounts of all essential nutrients *except* vitamin D, calcium, and phosphate, the raw materials important in bone mineral formation. Diet alone, however, will not cause rickets. Animals deprived of sunlight and fed a rachitogenic diet grow normally in all respects except that, because they lack bone mineral, they develop severe or “florid” rickets. But animals deprived of sunlight and fed a diet inadequate in many of the chief nutrients grow poorly, and the rickets they develop is difficult to discern against a background of other vitamin deficiencies.

By 1919, it was becoming clear that sunlight was the crucial factor in preventing rickets. Kurt Huld-schinsky (1919) cured the disease by exposing patients to a mercury-vapor lamp, thereby showing the importance of the ultraviolet (UV) portion of sunlight. In a stroke of genius, he also showed that irradiation of one arm of a rachitic child cured the skeleton throughout the body, *including the other arm*. He invoked the concept that vitamin D must be “a hormone” because it was a chemical that could heal at a distance. Only much later were Adolf Windaus and colleagues (1936) able to show that vitamin D is made in the skin from a precursor, 7-dehydrocholesterol, thereby completing our understanding of this aspect of vitamin D.

Meanwhile, in the 1920s, nutritional biochemists, including Mellanby and McCollum, were busy isolating several essential nutrients in foodstuffs. They too were able to cure rickets by administration of a fat-soluble substance, termed vitamin D by McCollum to distinguish it from vitamin A, which cured xerophthalmia or night blindness (Mellanby 1919; McCollum et al. 1922). In 1928, Windaus received the Nobel Prize principally for his elucidation of the structure of sterols, including vitamin D. Interestingly, Hess and Harry Steenbock separately had been able to produce vitamin D in food by irradiating it with ultraviolet light (Hess and Weinstock 1924; Steenbock and Black 1924). It became clear that the very process occurring in the skin could be mimicked in the test tube by subjecting certain plant oils or even yeast, containing plant sterols, to UV light. Thus vitamin D₂ was born, and with it, food fortification.

In more recent times, DeLuca, the last graduate student of Steenbock at the University of Wisconsin, showed that vitamin D is converted in the liver to a metabolite, 25-hydroxyvitamin D (Blunt, Schnoes, and DeLuca 1968). This is the main transport form of vitamin D in the body, and its blood level reflects the body's supply of vitamin D. Following this discovery, several groups, most notably those of Kodicek at Cambridge, DeLuca at Wisconsin, and Anthony Norman at the University of California, were able to demonstrate and identify the hormonally active form of vitamin D, calcitriol, for the first time (Fraser and Kodicek 1970; Myrtle, Haussler, and Norman 1970; Holick et al. 1971). Also known as 1,25-dihydroxyvitamin D, this hormone is made in the kidney, but it is important to note that its level in the blood is not simply a reflection of the exposure to sunlight or a measure of dietary intake. The human body can store the fat-soluble vitamin D precursor, make just enough calcitriol hormone for its needs, and save the rest of the vitamin D for "hard times." In the case of our ancestors who lived in the higher latitudes of northern Europe and were exposed to sunlight on a seasonal basis, these stores of vitamin D must have been crucial to help them get through each winter without the development of rickets or osteomalacia.

Foodstuffs and Vitamin D

As already pointed out, diet plays a secondary role in maintaining our supply of vitamin D. Nevertheless, diet can assume a critical importance when exposure to sunlight is compromised. Somewhat surprisingly, most foodstuffs are devoid of vitamin D. The only significant sources are animal liver (vitamin D stores of other vertebrates), egg yolks, and fish oils. Milk, generally thought of as a major source of vitamin D, is not rich in the vitamin, and human milk is an extremely poor source of it. It is a well-established observation in contemporary pediatric medicine that most cases of rickets in infants are found in those who are breast-fed, born in the fall, and kept out of the sun during winter months. Of course, the formula-fed infant of today receives milk fortified with vitamin D.

Most grains, meat, vegetables, and fruits are virtually devoid of measurable amounts of vitamin D, although experts in this field are still perplexed by how such creatures as nocturnal bats surviving on a diet of fruit and insects can avoid rickets! It is possible (although unsubstantiated) that exposure of certain foods (e.g., vegetables or fruits) to sun-drying may generate antirachitic activity, presumably because plant ergosterol would be converted into vitamin D₂. Some cultures (e.g., the Chinese) have a tradition of drying vegetables in the sun, which may increase their vitamin D content.

Recent reports indicate that infants fed a so-called macrobiotic diet, consisting of unpolished rice, pulses, and vegetables with a high fiber content along

with small additions of seaweeds, fermented foods, nuts, seeds, and fruits, are particularly susceptible to rickets. In one group of Caucasian children in the Netherlands (Dagnelie et al. 1990), 28 percent had physical symptoms of rickets in late summer, and this statistic rose to 55 percent by the following spring. (One might hope that those who extol this sort of diet will modify their teachings so that small amounts of fatty fish might be included, which would supply much-needed vitamin D.)

The association of vitamin D with fish oils, particularly fish-liver oils, is an interesting one recognized well before the formal discovery of vitamin D and predating even the discovery of the importance of sunlight. In 1789, a Manchester physician named Thomas Percival wrote about the medicinal uses of cod-liver oil at the Manchester Infirmary shortly after it had been introduced into British pharmacopoeia. In fact, one might argue that cod-liver oil is a medicine and not a food, but this is a fine point. There is, however, almost universal agreement that it is not a particularly good-tasting substance and hardly a favorite of children. Hess (1929) wrote that cod-liver oil's chief disadvantages lay in taste and odor. Moreover, it was not always completely prophylactic against rickets.

Hess's last comment is a reference to the variable potency of cod-liver oil because it is a natural product and thus subject to seasonal variation dependent upon the diet of the codfish. A more acceptable but less effective source of vitamin D is fish itself, particularly the fatty saltwater fish: herring, mackerel, tuna, halibut, and salmon. These fish have fat stores in the muscle, and because vitamin D is fat-soluble it is found throughout these fat deposits and is not confined to the liver, as in the cod. Fish roe, like eggs, also contain vitamin D. W. F. Loomis (1970), in a review of rickets, speculated that certain social practices, such as the Christian tradition of serving fish on Friday, might be adaptive responses to rickets. Another was June weddings, which tend to bring the first baby in the spring and permit the rapid growth phase of the first six months of life in summer sunshine. By contrast, the fall baby historically lacked vitamin D because of an infancy during the winter months.

Fortification of Food with Vitamin D

With the discovery that irradiation of ergosterol could produce a molecule (later identified as vitamin D₂) with potent antirachitic activity came the realization that such a preparation could be added to foodstuffs rendering low-vitamin D foods useful in the fight against rickets. As noted earlier, the two problems with foods containing vitamin D are that there are too few of them and that even those foods that contain vitamin D vary widely in potency. Steenbock had the idea to fortify staples of the diet such as breakfast cereals, milk, and margarine with vitamin D in the form of irradiated ergosterol, and because this supple-

ment has a narrower range of variability, the potency of such foods could be assured with some degree of confidence. Some U.S. states and Western countries fortify only milk or margarine; others include breakfast cereals as well. Nonetheless, fortifying even these few foods with vitamin D has virtually eradicated the incidence of rickets in the Western world. The fortification of foods with vitamins including vitamin D is arguably one of the most important medical achievements of the twentieth century and certainly a major achievement of the nutritional sciences.

Nowadays, pure crystalline vitamin D₃ is used in food fortification rather than Steenbock's irradiated ergosterol. Nevertheless, Steenbock's "invention" represents one of the earliest examples of a university (Wisconsin-Madison) patenting the application of fundamental scientific research in the biomedical field. The discovery led to the inception of a new university structure, WARF (Wisconsin Alumni Research Foundation), an institution designed to manage patentable research and recycle profits from such discoveries. Aside from serving as a model for many other similar institutions in the United States and around the world, WARF has spawned a number of other products from its profits on vitamin D fortification, including the rodent poison warfarin and most of the new metabolites of vitamin D itself, including calcitriol, identified and synthesized in the laboratory of DeLuca.

Despite the incredible success of food fortification in the eradication of rickets, over the years the process has met some resistance and even outright opposition amid fears that vitamin D in megadoses could cause hypercalcemia and, consequently, kidney damage. These fears are largely groundless because the doses required to produce renal damage are massive and could not be acquired by ingestion of large amounts of foods, even those fortified with vitamin D (cod-liver oil excepted). Yet one of the most notable examples of resistance came in the Province of Quebec, Canada. Health authorities in this province steadfastly resisted the fortification of dairy products with vitamin D until the early 1970s, when finally they bowed to pressure from a group headed by the notable clinical geneticist Charles Scriver to reduce rickets in the French-Canadian population. Following this decision, statistics from one Montreal hospital (Sainte-Justine pour les Enfants) showed a decline in the annual incidence of rickets from 130 per 1,000 to zero in an eight-year span between 1968 and 1976, which coincided with the introduction of provincial legislation making it mandatory for dairies to fortify milk (Delvin et al. 1978). Today, a similar low incidence of rickets can be documented in every children's hospital in North America. Gone too is the once familiar bowleggedness, the signature of rickets, which often remained with the victim for life and was so common in the Great Depression of the 1930s.

With the end of World War II came mandatory

rationing and governmental food fortification in Western Europe. But in 1957 politicians in the United Kingdom bowed to political pressure and drastically reduced vitamin D fortification of cereals and powdered milk following an "outbreak" of infantile hypercalcemia. At the time, the outbreak was believed to be caused by overfortification of food with vitamins D and A, although Donald Fraser, a noted Canadian pediatrician who researched the evidence at that time, now feels that this conclusion was probably incorrect (Fraser et al. 1966). The incidence of infantile hypercalcemia today is now no greater in countries that fortify than those that do not. Furthermore, the disease seems to result from an insult to the fetus in utero rather than a problem of the young child overindulging in fortified food, such as infant formula. Nevertheless, the condition results in mental retardation and heart problems that are largely irreversible and thus must be taken seriously.

To this day, the United Kingdom permits fortification of margarine only, probably at the expense of some increased incidence of rickets and osteomalacia in the population. Certainly, blood plasma levels of vitamin D and its metabolites (except calcitriol) are lower, presumably reflecting reduced stores of vitamin D in Britons when compared to a similar population of Americans and Canadians at the same latitude but given fortified food. In summary, it can be stated that the advantages of food fortification with vitamin D far outweigh any disadvantages. The main value of food fortification with vitamin D is to provide a *constant* year-round supply of the essential nutrient in the diet to augment the seasonal production in the skin.

Geographical Aspects of Vitamin D

Synthesis of vitamin D is both season- and latitude-dependent because vitamin D is made in the skin only by exposure to wavelengths in the UV spectrum of sunlight, and these wavelengths are absorbed by the ozone layer of the atmosphere. Only near the equator does the sunlight remain at an angle high enough for UV rays to penetrate the atmosphere on a year-round basis. Michael Holick of Boston City Hospital conducted experiments in which test tubes containing 7-dehydrocholesterol (the skin precursor to vitamin D) were exposed to light at various times of the day at latitudes from Caracas to Edmonton or Glasgow. He concluded (Webb, Kline, and Holick 1988) that in a city such as Boston, vitamin D is synthesized *only* in the months from April to October. Such a result implies that in most of the northern cities of the world, production of vitamin D₃ is seasonal and rickets might result at such latitudes if vitamin D stores are depleted *and* there is no dietary vitamin D.

Historical records also consistently reveal the geographical segregation of rickets to more northerly latitudes. August Hirsch (1883–6, vol. 3), for example, discussed the paucity of data on the geographical

distribution of rickets. Theobald Palm, a Western medical missionary to Japan and China, concluded in a valuable paper in 1890 that sunshine is the main etiological factor in rickets. He compiled anecdotal reports suggesting that rickets was rare in southern China but more common in northern regions of the same country. However, reports in the early 1900s suggested a similar incidence of rickets in New Orleans (30° latitude) and New York (40° latitude). Hess (1929) ascribed this similarity to equivalent average annual sunshine exposure (2,519 versus 2,557 hours). It is interesting to note that of the cities cited by Hess for the year 1923, Phoenix, Arizona, has the highest amount of annual sunshine with 3,752 hours (and no rickets in the population), whereas Glasgow (a hotbed of rickets) has the lowest with 1,086 hours. Since UV light penetration is greater at higher altitudes, we would expect the incidence of rickets to be low in mountain cities, and early studies in cities like Denver (5,000 ft. above sea level) bear this out. Even earlier, in 1844, a Swiss physician had pointed out that rickets was more prevalent in the lowlands than the mountains, again presumably reflecting the relative exposure to UV light (Hess 1929: 50). In more recent times, the pollution of major cities has modified local UV exposure and may have increased our predisposition to rickets by reducing skin synthesis of vitamin D. Thus, geography plays a major role in the distribution of rickets, but the relationship is not a simple one directly dependent on degrees of latitude. It is further modified by climate, altitude, and degree of pollution.

The geography of vitamin D and rickets is also modified by diet. It is important to note that many of the peoples who live in northern latitudes depend upon the sea for their survival (e.g., Eskimos, Haida Indians, Greenlanders, Scandinavians). It is tempting to speculate that the natural incidence of rickets in these cultures has been ameliorated by the higher vitamin D content of their diets (e.g., from fish oils). One anecdotal story relating to the higher fat-soluble vitamin content of Arctic meats is the experience of early Arctic explorers. Several were forced to shoot Arctic animals (such as polar bears) and eat almost all edible organs in order to survive. As a result, some contracted hypervitaminosis A, which causes disorientation and brain swelling, probably from ingestion of polar bear liver containing vast stores of vitamins A and D. Eskimos apparently avoid the liver of the polar bear but still get enough vitamins D and A from the rest of their diet.

Social and Ethnic Aspects of Vitamin D

Because vitamin D is produced in the skin by exposure to UV light, social practices relating to exposure of the skin (such as clothing habits or sunbathing) can be of paramount importance to this production. Primitive humans evolved near the equator and are popularly depicted wearing minimal clothing. These

considerations, plus the belief that primitive humans must have spent much of their lives outdoors, suggest that their contraction of rickets was unlikely. But as humans increased in numbers and moved into more and more inhospitable climes, the need for clothing would have become greater and the synthesis of vitamin D in the skin would have been compromised. It is possible that sun worship in early cultures reflects a realization of the importance to health of skin exposure to the sun.

Adaptation to sunlight exposure involved the development of melanin pigment in the skin. People exposed to maximal amounts of sunlight at the equator were the darkest, and those at the highest latitudes of Europe were the lightest. Although the subject is somewhat controversial, modern research indicates that skin melanin not only protects against the harmful effects of sunlight (e.g., skin cancer) but reduces the efficiency of the synthesis of vitamin D. Loomis (1970) suggests that summer bronzing in white populations is a seasonal adaptation to UV light exposure, though there is no evidence that vitamin D synthesis is regulated in an individual by such seasonal pigmentation. However, there is also no evidence that excessive exposure to sunlight can cause hypervitaminosis D and toxicity. The interrelationship of skin pigmentation and migration has some relevance to nutrition. Over the course of the past few centuries there has been considerable migration of certain peoples around the world so that cultural groups exist in climates that "their skin was not designed for."

Much has been written about whites developing skin cancer in Arizona and Australia, and considerable evidence is accumulating that indicates rickets and osteomalacia are common among Asian groups in England and Scotland (Felton and Stone 1966; Ford et al. 1973; Dent and Gupta 1975). There is some possibility that the problems of Asians in Britain are exacerbated by a high-phytate, high-phosphate diet (which tends to chelate calcium in the lumen of the intestine) combined with the avoidance of the typical vitamin D-fortified staples that whites consume.

Black children growing up in southern U.S. cities during the last part of the nineteenth century and the first decades of the twentieth century endured rickets almost as a rite of passage. And Rastafarians living outside of the UV climate of the West Indies may also be at risk for rickets. This is because of strict dietary practices that forbid consumption of artificial infant formulas, coupled with reduced opportunities for UV exposure in inner-city apartment complexes and a skin pigment unsuited to their "new" more northerly homes. There is also speculation (Ward et al. 1982) that the vitamin D-rich fish component of the Jamaican Rastafarian diet has been dropped by those living abroad. With the generation of multicultural societies such as currently exist in the United States, Canada, and many former colonial European coun-

tries, this potential problem of widespread rickets needs to be better understood and *dietary* solutions need to be formulated.

Clothing must have been developed for both staying warm in cooler climates and combating the harmful effects of UV light in warmer climates. Some cultures, particularly Moslem and Hindu groups, have very strict traditions of purdah, the practice of keeping the female skin out of the gaze of the public and therefore out of the sun. Though this strict policy does not apply to female children below the age of puberty, Saudi Arabian physicians report a higher female to male ratio in rickets patients that they ascribe to the higher UV exposure of the male children. In most cases, however, the practice of purdah cannot be directly blamed for childhood rickets in Saudi Arabia and North Africa, but it can be held responsible for osteomalacia observed in older females in such populations. And because these women with low reserves of vitamin D bear infants, the low UV exposure might be indirectly responsible for rickets. Detailed studies of the etiology of the rickets observed in Moslem groups (Belton 1986; Underwood and Margetts 1987; Elzouki et al. 1989) suggest the existence of maternal vitamin D deficiency, which results from the extra burden placed upon the mother to provide calcium for the skeleton of the growing fetus. Placental and lactational transfer of vitamin D stores from mother to the neonate are thus minimal, and such deficiencies, coupled with inadequate UV exposure, can result in infantile rickets. It is likely that the strict moral standards that prevailed in past centuries in western Europe leading to limited skin exposure also contributed to the well-documented higher incidence of osteomalacia in women.

Glenville Jones

Bibliography

- Belton, N. 1986. Rickets – not only the “English Disease.” *Acta Paediatrica Scandinavica* 323 (Suppl): 68–75.
- Blunt, J. W., H. K. Schnoes, and H. F. DeLuca. 1968. 25-Hydroxycholecalciferol: A biologically active metabolite of vitamin D. *Biochemistry* 7: 3317–22.
- Dagnelie, P., F. Vergote, W. van Staveren, et al. 1990. High prevalence of rickets in infants on macrobiotic diets. *American Journal of Clinical Nutrition* 51: 202–8.
- Delvin, E. E., F. H. Glorieux, M. Dussault, et al. 1978. Simultaneous measurement of 25-hydroxycholecalciferol and 25-hydroxyergocalciferol. *Medical Biology* 57: 165–70.
- Dent, C., and M. Gupta. 1975. Plasma 25-hydroxyvitamin-D levels during pregnancy in Caucasians and in vegetarian and non-vegetarian Asians. *Lancet* 2: 1057–60.
- Elzouki, A., T. Markstead, M. Elgarrah, et al. 1989. Serum concentrations of vitamin D metabolites in rachitic Libyan children. *Pediatric Gastroenterology and Nutrition* 9: 507–12.
- Felton, D. J. C., and W. D. Stone. 1966. Osteomalacia in Asian immigrants during pregnancy. *British Medical Journal* 1: 1521–2.
- Ford, J. A., D. C. Davidson, W. B. McIntosh, et al. 1973. Neonatal rickets in an Asian immigrant population. *British Medical Journal* 3: 211–12.
- Fraser, D., B. S. L. Kidd, S. W. Kooh, et al. 1966. A new look at infantile hypercalcemia. *The Pediatric Clinics of North America* 13: 503–25.
- Fraser, D. R., and E. Kodicek. 1970. Unique biosynthesis by kidney of a biologically active vitamin D metabolite. *Nature* 228: 764–6.
- Glisson, F. 1650. De Rachitide sive morbo puerili qui vulgo. *The rickets dicitur*. London.
- Hess, A. F. 1929. *Rickets including osteomalacia and tetany*. Philadelphia, Pa.
- Hess, A. F., and M. Weinstock. 1924. Antirachitic properties imparted to inert fluids and to green vegetables by ultra-violet irradiation. *Journal of Biological Chemistry* 62: 301–13.
- Hirsch, A. 1883–6. *Handbook of geographical and historical pathology*, 3 vols., trans. Charles Creighton. London.
- Holick, M. F., H. K. Schnoes, H. F. DeLuca, et al. 1971. Isolation and identification of 1,25-dihydroxycholecalciferol. A metabolite of vitamin D active in intestine. *Biochemistry* 10: 2799–804.
- Huldschinsky, K. 1919. Heilung von Rachitis durch künstliche Höhensonne. *Deutsche Medizinische Wochenschrift* 45: 712–13.
- Loomis, W. F. 1970. Rickets. *Scientific American* 223: 76–91.
- McCollum, E. V., N. Simmonds, J. E. Becker, et al. 1922. Studies on experimental rickets. XXI. An experimental demonstration of the existence of a vitamin which promotes calcium deposition. *Journal of Biological Chemistry* 53: 293–312.
- Mellanby, E. 1919. An experimental investigation on rickets. *Lancet* 1: 407–12.
- Myrtle, J. F., M. R. Haussler, and A. W. Norman. 1970. Evidence for the biologically active form of cholecalciferol in the intestine. *Journal of Biological Chemistry* 245: 1190–6.
- Palm, T. A. 1890. The geographical distribution and etiology of rickets. *The Practitioner* 45: 270–9.
- Percival, T. 1789. *Essays medical, philosophical and experimental on the medical use of cod-liver oil*, Vol. 2. London.
- Raczynski, J. 1913. Recherches expérimentales sur la manque d'action du soleil comme cause de rachitisme. *Comptes rendues de l'association internationale de pediatrics* 308. Paris.
- Shorter, E. 1982. *A history of women's bodies*. New York.
- Steenbock, H., and A. Black. 1924. The induction of growth-promoting and calcifying properties in a ration by exposure to ultraviolet light. *Journal of Biological Chemistry* 61: 408–22.
- Underwood, P., and B. Margetts. 1987. High levels of childhood rickets in rural North Yemen. *Society of Scientific Medicine* 24: 37–41.
- Ward, P., J. Drakeford, J. Milton, and J. James. 1982. Nutritional rickets in Rastafarian children. *British Medical Journal* 285: 1242–3.
- Webb, A. R., L. Kline, and M. F. Holick. 1988. Influence of season and latitude on the cutaneous synthesis of vitamin D₃. *Journal of Clinical Endocrinology and Metabolism* 67: 373–8.
- Whistler, Daniel. 1645. De morbo puerili Anglorum, quem patrio idiomate indigenae vocant The Rickets. M.D. thesis, University of Leiden. Leiden, Netherlands.
- Windaus, A., F. Schenck, and F. van Werder. 1936. Über das antirachitisch wirksame Bestrahlungsprodukt aus 7-Dehydrocholesteria. *Hoppe Seyler's Zeitschrift für Physiologische Chemie* 241: 100–3.

IV.A.5 Vitamin E

As any nutritional text dated prior to 1970 will indicate, vitamin E has not received much respect from nutritionists. In such texts it is often placed after vitamin K, in the miscellaneous category. This is because it took a good 40 years from its discovery in 1923 (Evans and Bishop 1923) to demonstrate a clear-cut human deficiency disease for vitamin E. Though numerous studies had shown vitamin E to be an essential component of animal diets, deficiency symptoms varied from one species to the next, from reproductive disorders in rats to vascular abnormalities in chickens. Thus, it was not clear that humans had an obligatory requirement for vitamin E. Recent research, however, has shown that this is indeed the case and that vitamin E is just as important to human nutrition as the other vitamins. It is, therefore, pleasing to see that vitamin E is now placed in its proper place in the alphabet of vitamins.

Vitamin E is the nutritional term used to describe two families of four naturally occurring compounds each, the tocopherols and the tocotrienols (Pennock, Hemming, and Kerr 1964). Tocopherols and tocotrienols both contain a chroman ring, which is essential for biological activity, but differ in the degree of saturation of their fatty side chains. They are otherwise interchangeable in their biological role. Each family comprises alpha, beta, gamma, and delta forms, which differ significantly in their potency. Thus, alpha-tocopherol represents the principal source of vitamin E found in the human diet with a small contribution also coming from gamma-tocopherol (Bieri and Evarts 1973). Many texts, including this one, use the terms alpha-tocopherol and vitamin E interchangeably.

Located in the cellular membranes, alpha-tocopherol helps the cell to resist damage from powerful oxidants known as free radicals. These are generated naturally inside the body as by-products of fuel oxidation, and they can be generated artificially by external factors such as radiation, chemotherapy, or pollutants. B. Ames (1983) believes that aerobic respiration using oxygen is the most important of these external factors in generating free radicals and that antioxidants such as vitamin E help to resist free-radical damage. Because of its stabilized chroman ring structure, vitamin E is able to mop up these free radicals and their immediate products, minimizing the damage to the lipids of the membrane and therefore maintaining cell membrane integrity. Simply put, vitamin E stops the fat of the body from turning rancid.

History of the Discovery of Vitamin E

The decade between 1915 and 1925 was one of the most productive in the history of nutrition. The use of semipurified or semisynthetic diets allowed researchers to demonstrate the essential nature of individual components of the diet and, more specifi-

cally, to recognize the existence of the vitamins A, B₁, C, and D. Nutritionists (Osbourne and Mendel 1919; Mattill and Conklin 1920) showed that though these diets were able to maintain life, they failed to support reproduction in laboratory animals. Reproductive biologists, thus, realized that the estrous cycle of the female rat or testicular development in the male constituted useful animal models for studying the essential nature of nutritional factors (Long and Evans 1922). Using a basal semipurified diet comprising casein, starch, lard, butter, fats, and brewer's yeast, which would allow rats to grow but not reproduce, groups headed by Herbert M. Evans and L. B. Mendel set about the laborious task of finding a substance that would promote fertility.

Detailed historical accounts of the events surrounding the discovery of vitamin E, its chemistry and biology, have been published by two pioneers in the field, Evans (Evans 1962) and Karl E. Mason, a student of Mendel (Mason 1977). Their recollections are concise, modest accounts of the important milestones in the field, and unlike other accounts of historical events from rival camps, the two actually agree! The reader is referred to the often flowery account of Evans, who describes the identification of a factor in lettuce and wheat germ required for preventing resorption of fetuses in the pregnant rat. He writes:

Good fairies attended every phase of the advent and early history of vitamin E. We turned our attention at once to the prevention rather than alleviation of these strange resorptions – a prevention which might disclose at once what individual natural foodstuffs carried a missing needed substance. Lettuce, relished by these poor sufferers of our rancid lard diet, was spectacularly successful, and we may have entertained the conviction that vitamin C which was not essential for growth was necessary in pregnancy, had we not quickly shown that not the aqueous, but only the fatty, component of these leaves, the chlorophyll-rich green oil, had worked the good result. Then, to our surprise, wheat was equally remedial, and the concept that vitamin C was involved could not, of course, survive. The good fairies accompanied me to the large Sperry flour mill at a neighboring town, Vallejo, where I found three great streams flowing from the milling of the wheat berry: the first constituted the outer cover or chaff; the second the endosperm, the white so-called flour; and the third which came in flattened flakes, stuck into such units by its oil content – the germ. Night had not fallen that day, before all these components were fed to groups of carefully prepared females – animals which had begun gestation on vitamin E-low diets and were fed both the watery and fatty solutions. Single daily drops of the golden yellow wheat

germ oil were remedial. That an oil might enrich the embryo's dietary needs for vitamin A and vitamin D, the only fat-soluble vitamins then known, was negated at once when we added the well-known rich source of vitamins A and D, cod liver oil, an addition which did not lessen but increased and made invariable our malady. (Evans 1962: 382)

From the words Evans used to describe his rats, we can clearly discern that he cared about the animals he used in those early studies. The factor initially described as "antisterility factor X" was born (Evans and Bishop 1922). Parallel studies by Mason (1925), working with the male rat, showed that the same factor appeared essential to prevent testicular degeneration in the rat. The name, vitamin E, seems to have been suggested by an opportunist, Barnett Sure, in 1924, based upon the vitamin nomenclature of the time, although the endorsement of this title by Evans helped to gain it widespread use (Evans 1925).

Following the discovery of vitamin E, nutritionists spent the next decade describing symptoms of its deficiency in a variety of animals, *except* humans. Research also focused on its chemical nature. Finally in 1936, Evans, working with Gladys and Oliver Emerson, published the identity of vitamin E from wheat germ oil as an alcohol with the chemical formula ($C_{29}H_{50}O_2$). Evans named this substance "alpha-tocopherol." The origin of this name is described in his personal account:

I well remember their (the Emersons') plea to me to suggest a proper name for their purified substance when success crowned their efforts. I promptly invited George M. Calhoun, our professor of Greek to luncheon in Berkeley in our small Faculty Club. "Most scientists, medical men especially," said Calhoun, "have been guilty of coining Greek-Latin terms, bastards, of course, and we might have to do this." "What does the substance do?" he asked. "It permits an animal to bear offspring," I replied. "Well, childbirth in Greek is *tocos*," he said, "and if it confers or brings childbirth, we will next employ the Greek verb *phero*. You have also said that the term must have an ending consonant with its chemical - 'ol', it being an alcohol; your substance is 'tocopherol,' and the pleasant task assigned me quickly solved and not worth the delightful four-course dinner you have arranged." (Evans 1962: 383)

It is, therefore, evident that vitamin E was renamed alpha-tocopherol for the meager price of a four-course meal.

Erhard Fernholz provided the structural formula for alpha-tocopherol in 1938, and it was first synthesized chemically by the brilliant Swiss chemist Paul Karrer working at Hoffmann-LaRoche laboratories in Basel (Karrer et al. 1938).

Vitamin E Deficiency in Humans

During the period 1925 to 1935 there appeared an ever increasing and confusing series of papers in the nutritional literature showing the importance of vitamin E in the prevention of reproductive defects in rats; prevention of embryonic mortality and encephalomalacia in chicks; and prevention of nutritional muscular dystrophy in guinea pigs and rabbits. How could deficiency of a single nutritional factor cause all these different pathological changes in different laboratory animals but have no apparent parallel condition in humans? Even in avian species like the chick and duck the symptoms were not consistent.

Vitamin E deficiency appeared in different sites: muscular dystrophy in ducklings and vascular changes and encephalopathy in chicks. The common thread tying together all these apparently different defects ascribed to vitamin E deficiency was not immediately evident. Although H. S. Olcott and H. A. Mattill (1931) had pointed out the association of antioxidants and vitamin E in lettuce oil, it was not until 1937 after the discovery of alpha-tocopherol that it became clear that alpha-tocopherol was an antioxidant (Olcott and Emerson 1937). It was still later through the work of researchers such as A. L. Tappel and J. G. Bieri in the late 1950s and early 1960s, that the free-radical theory of lipid peroxidation and the function of vitamin E as an antioxidant developed to the point where the conflicting findings could be rationalized using a single mechanism (Tappel 1962).

According to this free-radical theory, membrane lipids in all membranes of the body should be susceptible to oxidative damage, and the exact location of damage observed in different animal species would depend upon variables such as specific membrane lipid composition, vitamin E content, and cellular metabolic rate. Based upon this theory it became easy to rationalize various patterns of damage due to vitamin E deficiency in different animal species. With this new understanding of the nature of vitamin E action came the realization that vitamin E deficiency in humans might show up in different sites (i.e., different membranes) from those showing vitamin E deficiency in animal tissues. Moreover, the recognition that the common feature of all these deficiency symptoms was damage caused by free radicals began a search for human individuals who might generate or be exposed to higher levels of such destructive factors and who might, therefore, develop vitamin E deficiency. These individuals turned out to be prematurely born infants.

Looking back at the extensive literature on vitamin E in humans, it is clear that clues of vitamin E deficiency in pediatric medicine were emerging even in the late 1940s. The advent of modern neonatal units led to the use of respirators in which newborns were kept in an oxygen-rich environment. Premature infants were particularly favored for this treatment, and two conditions suggesting vitamin E deficiency

affecting premature infants came to light: hemolytic anemia, in which erythrocyte membranes have an increased tendency to rupture; and retrolental fibroplasia, in which the blood vessels of the developing retina are damaged, leading to scarring and blindness (Farrell 1980). First observed in the 1940s, these conditions were documented and rationalized for the first time with the advent of the free-radical theory.

Free radicals, generated from the oxygen-rich gaseous mix used in such respirators, attacked cellular membranes, in particular those of red blood cells and the cells of the retina. It should be noted that although premature infants are susceptible to damage resulting from vitamin E deficiency, this is due only in part to an increased "insult" from oxidants; it is also partly the result of reduced vitamin E stores and lower vitamin E levels in their blood. Not surprisingly, the severity of retrolental fibroplasia is markedly decreased by pharmacological doses of vitamin E (Hittner et al. 1981). As the assault on cellular membranes in such fast-growing children is diminished and as vitamin E stores are bolstered in the first weeks of life, the risk of vitamin E deficiency decreases substantially. Consequently, vitamin E deficiency is very difficult to demonstrate in healthy children or in adults eating a standard balanced diet.

Toward the end of the 1950s, the Food and Nutrition Board of the National Research Council of the United States funded a long-term, six-year, dietary study of the relationship between blood vitamin E levels and consumption of polyunsaturated fatty acids (PUFA) in a group of male subjects who were given a diet low in vitamin E. The study became known as the Elgin Project, after the hospital in Illinois where it was carried out (Horwitt 1960). Institutionalized subjects were closely monitored for symptoms of vitamin E deficiency while they underwent the increased stress of a diet with a higher PUFA: vitamin E ratio. Not surprisingly, the subjects failed to develop a clear-cut clinical syndrome except that plasma vitamin E levels dropped and red cells showed an increased tendency to hemolysis, but it took two years for these changes to occur.

If vitamin E deficiency does occur in humans beyond the neonatal period, it is usually secondary to another disease state that results in malabsorption of vitamin E from the diet. Since vitamin E is absorbed with the fat in the diet, conditions that result in fat malabsorption, such as chronic cholestasis (liver disease) or cystic fibrosis where there is pancreatic blockage, will result in vitamin E deficiency. Another rare cause of vitamin E deficiency is the genetic absence of beta-lipoprotein, the transporter of vitamin E in the blood, in a condition known as abetalipoproteinemia (Muller, Lloyd, and Bird 1977).

In such individuals with vitamin E supply or transport problems, there is an increased tendency of red cells to hemolyze and also a shortened red cell survival time. More recently, the medical fraternity has

begun to recognize muscle and neurological problems that result from human vitamin E deficiency (Sokol 1988). These abnormalities include loss of reflexes and gait disturbances that are accompanied by pathological changes. Autopsies performed on patients with cystic fibrosis show more advanced axonal degeneration than would be expected of normal age-matched individuals (Sung 1964). It would be surprising if humans were spared all of the neurological and muscular defects observed in animals. However, such human vitamin E deficiency is rarely accompanied by the acute muscle, brain, and blood vessel defects observed in laboratory animals fed semisynthetic diets. We should be thankful that it is not.

Dietary Sources of Vitamin E

The substance alpha-tocopherol is present in the diet in a variety of plant oils, including wheat germ oil and the lettuce oil discovered in pioneering work. Soya bean and other plant oils, which became more common dietary components in the latter half of the twentieth century, contain gamma-tocopherol, a less potent version of vitamin E. It is interesting to note that the unsaturated fatty acids, which are present in plant oils and are so prone to free-radical oxidation in our food, are accompanied by the highest concentrations of vitamin E. It is as if nature realized their sensitivity and put in a natural preservative to protect these oils (Bieri, Corash, and Hubbard 1983). Animal fats contain lower amounts of vitamin E, and fish oils have variable amounts depending upon the diet of the fish and the age of the fish oil. Though cod-liver oil has been shown to contain vitamin E, it also contains a high concentration of polyunsaturated fatty acids that can easily become oxidized by free radicals and, consequently, as noted by Evans (1962), be deleterious rather than protective.

One special food associated with vitamin E is wheat germ bread. Sometimes referred to in Europe as Hovis (from the Latin *hominis vis* meaning "the strength of man"), this bread is made with flour containing five times as much of the fatty germ as whole-meal bread. The procedure by which the wheat germ is stabilized involves separating it from the flour, lightly cooking it in steam, and returning it to the flour. This process was jointly patented in the United Kingdom in 1887 by Richard "Stoney" Smith of Stone, Staffordshire, and Macclesfield Miller and Thomas Fitton of the firm S. Fitton and Son. This was about 35 years before the experiments of Evans, which led to the discovery of vitamin E. Although the health-promoting properties of Hovis were clearly recognized in early advertising campaigns, there is no evidence that Hovis was aimed at special groups of people (i.e., those with muscular, reproductive, or vascular problems). With such rich sources of vitamin E as Hovis bread in our normal diet, it is difficult to consume a diet that will result in vitamin E deficiency.

Other Uses of Vitamin E

The history of vitamin E is much shorter than that of the other vitamins, but the absence for so long of a clear-cut human deficiency syndrome allowed for the development of a number of bogus or exaggerated claims for vitamin E, or at least claims that have never been properly substantiated. Most of these have been built upon deficiency symptoms observed in laboratory animals and not in humans. One, based upon findings in the rat, is that vitamin E is a fertility vitamin. As a result, some physicians used the vitamin to treat a number of conditions that produced spontaneous abortions. In a series of poorly controlled, anecdotal studies, Evan Shute (1939) claimed that vitamin E was particularly effective in countering habitual abortion, which is defined as spontaneous abortion before the sixteenth week of gestation during three successive pregnancies. When others tried to confirm these findings, they failed to do so and were left to conclude, as did C.T. Javert, W.F. Finn, and H.J. Stander:

There is such a maze of literature the proper cognizance cannot be taken of all the pertinent articles. As the reader reviews them in order to develop his own philosophy let him [all obstetricians were male in those days] be reminded of the following important matters:

- i. the high percentage of success irrespective of which vitamin, hormone or method is employed;
- ii. the lack of specific information as to the pathogenesis of human spontaneous abortion. (1949: 887)

This skepticism was found to be well grounded; others who have reviewed three decades of experimentation on abortion have also concluded that vitamin E supplementation is ineffective (Marks 1962).

In addition, there have been numerous claims that large doses of vitamin E are beneficial in cardiovascular conditions such as angina, congestive heart failure, and peripheral vascular disease. Current medical opinion is that these claims are unproven. However, this debate was being revisited in the 1990s with the suggestion that the “oxidized” lipoprotein, LDL, may be the chief villain in atherosclerosis, and that its level in the bloodstream inversely correlates with alpha-tocopherol levels. New clinical studies of vitamin E and the risk of coronary heart disease in both male and female health workers have, again, suggested a protective role for the vitamin (Rim et al. 1993; Stampfer et al. 1993). Consequently, long-term supplementation with vitamin E to prevent heart disease is again being discussed.

The association of vitamin E deficiency with nutritional muscular dystrophy in guinea pigs and rabbits led to trials of its use as a supplement in Duchenne’s muscular dystrophy (Berneske et al. 1960). Because this condition is a genetically inherited disease with a

different etiology than the nutritional version, it is not surprising that vitamin E was ineffective, but this is not to say that vitamin E supplementation cannot help to alleviate some of the consequences of muscular and neurological disease and, therefore, benefit the patient. The recent elucidation of the molecular defect in Lou Gehrig’s disease (ALS) as a lack of the enzyme superoxide dismutase (Rosen et al. 1993) has already led to a reevaluation of dietary antioxidant supplementation in this disease because the enzyme works in concert with vitamin E to reduce free radicals inside the cell. It is hoped that the results will be more positive than those reported recently in another neurological condition, Parkinsonism, where alpha-tocopherol supplementation was unsuccessful as an adjunct therapy to the drug deprenyl (Shoulson et al. 1993).

The old adage that if a little is good, a lot must be better, pervades nutritional science. The association of vitamin E with muscle health has led to the use of vitamin supplements in sports (Cureton 1954). Consequently, vitamin E can be considered one of the earliest performance-enhancing drugs, in use long before anabolic steroids and growth hormones. However, a controlled study of Scottish track-and-field athletes in the early 1970s by I. M. Sharman, M. G. Down, and R. N. Sen dispelled the notion that vitamin E improved performance. Presumably the athletes agreed because they moved on to other treatments.

In the 1970s, megadose vitamin therapy came into vogue in North America, and Senator William Proxmire from Wisconsin, who advocated such self-treatments for improved health, subsequently led opposition to legislation that might have blocked such potentially dangerous megadose vitamin products. Simultaneously in vogue was the free-radical theory of lipid peroxidation, which resulted in the idea that the natural aging process is an inability of the organism to keep up with oxidation (Harman 1956).

A possible solution to the problem came from Ames (1983), who suggested that a battery of dietary and endogenously produced antioxidants, headed by vitamin E and also including glutathione, vitamin C, beta-carotene, and possibly uric acid, might help to slow down this process. Ames concluded that the natural aging process results from an inability to completely prevent the harmful effects of oxidation. As a consequence, megadoses of vitamin E have been consumed by thousands of North Americans in the hope that augmenting the antioxidant supply might help defeat free radicals and prevent aging. Although it is difficult to believe that supplemental dietary vitamin E will prevent aging if alpha-tocopherol levels are already adequate in the blood of the majority of adults, the toxicity symptoms (gastrointestinal disturbances) of vitamin E are minor. Thus there seems not much reason to prevent a little self-experimentation, although it is worth noting that this self-experimentation is ably assisted by the pharmaceutical industry.

As the antioxidant properties of vitamin E have

become more and more evident, the food and cosmetic industry have found wider and wider uses for it. Vitamin E is a poor antioxidant in vitro and is largely replaced by other substances when a preservative is needed in food. However, the vitamin is now added to a broad range of soaps and shampoos and other cosmetic products in the hope of convincing the buyer that it has special properties in skin and hair. The fact that human vitamin E deficiency is not accompanied by skin and hair problems, or that vitamin E might not even enter the hair or skin when applied via this route, does not seem to be a weighty argument for the consumer. Perhaps the most outrageous claim this contributor has come across in recent years is one stating that a cream containing vitamin E might protect against frostbite. It was applied by a Canadian sled-dog racer to the underside of her dogs' testicles to protect them from the harmful effects of the snow and cold (Sokol 1985).

The past two to three decades have, however, also seen some legitimate uses of vitamin E in medical science. The development of total parenteral feeding solutions in the 1970s included a recognition of the essential nature of vitamin E (Bieri et al. 1983). Long-term artificial feeding solutions are, therefore, supplemented with vitamin E. The widespread use of cancer chemotherapeutic agents such as adriamycin, which generates free radicals to kill cancer cells, can be augmented by vitamin E therapy to help protect surviving normal cells. Tissue transplantation normally involves maintenance of donor organs in oxygenated perfusate prior to surgery, a process that leads to free-radical generation. Use of antioxidants and free-radical scavengers helps improve the survival of such organs, presumably by minimizing the damage caused by free radicals. Children with cholestasis and poor vitamin nutrition receiving liver transplants are given water-soluble vitamin E preparations to help the donated organ survive. Vitamin E can also attenuate the damage caused by free radicals released during myocardial infarctions (Massey and Burton 1989). It appears that uses for vitamin E will continue to proliferate well into the new millennium.

Glenville Jones

Bibliography

- Ames, B. 1983. Dietary carcinogens and anticarcinogens. *Science* 221: 1256-64.
- Berneske, G. M., A. R. C. Butson, E. N. Gould, and D. Levy. 1960. Clinical trial of high dosage vitamin E in human muscular dystrophy. *Canadian Medical Association Journal* 82: 418-21.
- Bieri, J. G., L. Corash, and V. S. Hubbard. 1983. Medical uses of vitamin E. *New England Journal of Medicine* 308: 1063-71.
- Bieri, J. G., and R. P. Everts. 1973. Tocopherols and fatty acids in American diets. *Journal of the American Dietetic Association* 62: 147-51.

- Cureton, T. K. 1954. Effect of wheat germ oil and vitamin E on normal human subjects in physical training programs. *American Journal of Physiology* 179: 628.
- Evans, H. M. 1925. Invariable occurrence of male sterility with dietaries lacking fat soluble vitamin E. *Proceedings from the National Academy of Sciences USA* 11: 373-7.
1962. The pioneer history of vitamin E. *Vitamins and Hormones* 20: 379-87.
- Evans, H. M., and K. S. Bishop. 1922. The existence of a hitherto unrecognized dietary factor essential for reproduction. *Science* 56: 650-1.
1923. The production of sterility with nutritional regimes adequate for growth and its cure with other foodstuffs. *Journal of Metabolic Research* 3: 233-316.
- Evans, H. M., O. H. Emerson, and G. A. Emerson. 1936. The isolation from wheat germ oil of an alcohol, α -tocopherol having the properties of vitamin E. *Journal of Biological Chemistry* 113: 319-32.
- Farrell, P. M. 1980. Deficiency states, pharmacological effects and nutrient requirements. In *Vitamin E: A comprehensive treatise*, ed. L. J. Machlin, 520-600. New York.
- Fernholz, E. 1938. On the constitution of α -tocopherol. *Journal of the American Chemical Society* 60: 700-5.
- Harman, D. 1956. Aging: A theory based upon free-radical and radiation chemistry. *Journal of Gerontology* 11: 298-300.
- Hittner, H. M., L. B. Godio, A. J. Rudolph, et al. 1981. Retrolental fibroplasia: Efficacy of vitamin E in a double blind clinical study of preterm infants. *New England Journal of Medicine* 305: 1365-71.
- Horwitt, M. K. 1960. Vitamin E and lipid metabolism in man. *American Journal of Clinical Nutrition* 8: 451-61.
- Javert, C. T., W. F. Finn, and H. J. Stander. 1949. Primary and secondary spontaneous habitual abortion. *American Journal of Obstetrics and Gynecology* 57: 878-89.
- Karrer, P., H. Fritzsche, B. H. Ringier, and A. Salomon. 1938. Synthese des α -tocopherols. *Helvetica Chimica Acta* 21: 820-5.
- Long, J. A., and H. M. Evans. 1922. The oestrus cycle in the rat. *Memoirs of the University of California* 6: 1-51.
- Marks, J. 1962. Critical appraisal of the therapeutic value of α -tocopherol. *Vitamins and Hormones* 20: 573-98.
- Mason, K. E. 1925. A histological study of sterility in the albino rat due to a dietary deficiency. *Proceedings of the National Academy of Sciences USA* 11: 377-82.
1977. The first two decades of vitamin E. *Federation Proceedings* 36: 1906-10.
- Massey, K. D., and K. P. Burton. 1989. α -Tocopherol attenuates myocardial-membrane related alterations resulting from ischemia and reperfusion. *American Journal of Physiology* 256: H1192-9.
- Mattill, H. A., and R. E. Conklin. 1920. The nutritive properties of milk; with special reference to reproduction in the albino rat. *Journal of Biological Chemistry* 44: 137-57.
- Muller, D. P. R., J. K. Lloyd, and A. C. Bird. 1977. Long-term management of abetalipoproteinemia: Possible role for vitamin E. *Archives of Diseases of Childhood* 52: 209-14.
- Olcott, H. S., and O. H. Emerson. 1937. Antioxidants and the autoxidation of fats. IX. The antioxidant properties of the tocopherols. *Journal of the American Chemical Society* 59: 1008-9.
- Olcott, H. S., and H. A. Mattill. 1931. The unsaponifiable lipids of lettuce. II. Fractionation. *Journal of Biological Chemistry* 93: 59-64.
- Osbourne, J. B., and L. B. Mendel. 1919. The nutritive values of yeast protein. *Journal of Biological Chemistry* 38: 223-7.
- Pennock, J. F., F. W. Hemming, and J. D. Kerr. 1964. A reassessment of tocopherol chemistry. *Biochemical and Biophysical Research Communications* 17: 542-8.

- Rim, E. B., M. J. Stampfer, A. Ascherio, et al. 1993. Vitamin E consumption and the risk of coronary disease in men. *New England Journal of Medicine* 328: 1450–6.
- Rosen, D. R., T. Siddique, D. Patterson, et al. 1993. Mutations in Cu/Zn superoxide dismutase gene are associated with familial amyotrophic lateral sclerosis. *Nature* 362: 59–62.
- Sharman, I. M., M. G. Down, and R. N. Sen. 1971. The effects of vitamin E and training on physiological function and athletic performance in adolescent swimmers. *British Journal of Nutrition* 26: 265–76.
- Shoulson, I., S. Fahn, D. Oakes, et al. 1993. Effects of tocopherol and deprenyl on the progression of disability in early Parkinson's disease. *New England Journal of Medicine* 328: 176–83.
- Shute, E. V. 1939. The diagnosis and treatment of vitamin E deficiency. In *Vitamin E. A symposium held under the auspices of the Society of Chemical Industry*, ed. A. L. Bacharach and J. C. Drummond, 67–76. Cambridge.
- Sokol, A. 1985. Husky hounds are hard to beat in sled racing. *The Toronto Star*, Feb. 28, p. H8.
- Sokol, R. J. 1988. Vitamin E deficiency and neurological disease. *Annual Reviews of Nutrition* 8: 351–73.
- Stampfer, M. J., C. H. Hennekens, J. E. Manson, et al. 1993. Vitamin E consumption and the risk of coronary disease in women. *New England Journal of Medicine* 328: 1444–9.
- Sung, J. H. 1964. Neuroaxonal dystrophy in mucoviscidosis. *Journal of Neuropathology and Experimental Neurology* 23: 567–83.
- Sure, B. 1924. Dietary requirements for reproduction. II. The existence of a specific vitamin for reproduction. *Journal of Biological Chemistry* 58: 693–709.
- Tappel, A. L. 1962. Vitamin E as the biological lipid antioxidant. *Vitamins and Hormones* 20: 493–510.

cluded that the protective agent in the cabbage was vitamin C. However, when pure vitamin C became available, it was injected into the chicks and failed to prevent the hemorrhagia (Dam 1964).

Large doses of vitamins A and D (fed in the form of fish-liver oils) and commercial carotene also did not prevent the hemorrhagia (Dam 1935). Cereals and seeds did prevent the symptom, and green leaves and hog liver were found to be potent sources of the anti-hemorrhagic factor. Research groups led by Dam and H. J. Almquist worked independently to show that the factor was a new fat-soluble vitamin. Dam's report (Dam 1935) was followed in the same year by that of Almquist and E. L. R. Stokstad (1935). The factor was designated vitamin K (Dam 1935, 1964). According to Dam (1964: 10), "The letter K was the first one in the alphabet which had not, with more or less justification, been used to designate other vitamins, and it also happened to be the first letter in the word 'koagulation' according to the Scandinavian and German spelling."

The elucidation of the structure of vitamin K was achieved in a relatively short time because of the number of large research groups involved. Several reviews are available on the isolation and characterization of vitamin K (Doisy, Binkley, and Thayer 1941; Dam 1942, 1964; Almquist 1979). In 1943, Dam received the Nobel Prize for Physiology or Medicine for his discovery of vitamin K. It is interesting to note that Almquist, who published later but in the same year as Dam, did not share the Nobel Prize. Dam did share the prize with E. A. Doisy, who was honored for the isolation, chemical synthesis, and structural identification of one of the forms of vitamin K.

IV.A.6 Vitamin K and Vitamin K-Dependent Proteins

The term "vitamin K" was first introduced by Henrik Dam in 1935, following discovery of a fat-soluble substance that could prevent bleeding (Dam 1935, 1964). During the years 1928 to 1930, Dam conducted studies on the cholesterol metabolism of chicks at the University of Copenhagen. The chicks were being fed an artificial, practically sterol-free diet to which the then known vitamins A and D were added. He observed hemorrhages (bleeding) in different parts of the body in some of the chicks that had been on the diet for more than two or three weeks. In addition, blood that was examined from some of these chicks showed delayed coagulation or clotting (Dam 1964). The low amounts of cholesterol and fat in the diet were ruled out as the causes of the symptom.

Similar observations were made by W. D. McFarlane, W. R. Graham, Jr., and G. E. Hall (1931). W. F. Holst and E. R. Halbrook (1933) observed that the symptoms resembled scurvy, a disease caused by vitamin C deficiency, and could be prevented by addition of fresh cabbage to the diet. These investigators con-

The Structure of Vitamin K

Many compounds, all closely related, are now recognized as having antihemorrhagic activity (Table IV.A.6.1). Common to all is the methylnaphthoquinone nucleus. This nucleus, known as menadione, has not been isolated from natural sources, but it has been synthesized and does possess biological activity. Attached to the methylnaphthoquinone nucleus are carbon chains that vary both in nature and length. Based on the nature of these carbon side chains, the natural K vitamins are subdivided into two groups: phyloquinone (vitamin K₁) and the menaquinones (MK-*n*). Phyloquinone, the plant product, has a 20-carbon side chain with only one double bond. The menaquinones, synthesized by microorganisms, contain side chains with varying numbers of 5-carbon (isoprene) units but always with one double bond per 5-carbon unit. Chain lengths of the known menaquinones vary from 2 to 13 isoprene units. The *n* in MK-*n* refers to the number of 5-carbon units. A menaquinone with 4 isoprene units (20 carbons) in its side chain, for example, would be called menaquinone-4, abbreviated MK-4. Humans convert the synthetic menadione nucleus to menaquinone-4.

Table IV.A.6.1. *Vitamin K analogues*

Analogue	Symbol	Source
Menadione	K ₃	Synthetic
Phylloquinone	K ₁	Plants, animal tissues
Menaquinones	MK-2 to MK-13	Bacteria, animal tissues

Nutritional Aspects of Vitamin K

Since the human body cannot synthesize the naphthoquinone entity, all the K vitamins in humans, at least the quinone nucleus of these compounds, are of

extraneous origin. The K vitamins are synthesized by plants as phylloquinone and by bacteria in the intestines as the menaquinones (Table IV.A.6.1).

Phylloquinone, the form of vitamin K synthesized by plants, is the major dietary form of vitamin K (Booth et al. 1993; Booth, Sadowski, and Pennington 1995; Shearer, Bach, and Kohlmeier 1996). It is found in plants and in the tissues of plant-eating animals (Table IV.A.6.2) and is the form of vitamin K that is added to infant formulas and given to infants at birth.

The best sources of phylloquinone are green leafy vegetables and certain oils, such as soybean, rapeseed, and olive oil (Shearer et al. 1996). The values in

Table IV.A.6.2. *Phylloquinone content of common foods*

Phylloquinone concentration ranges in µg of phylloquinone per 100 g (3.5 oz.) food			
0.1–1.0	1–10	10–100	100–1,000
Avocado (1.0)	Apples (6)	Beans, runner (26)	Broccoli tops (179)
Bananas (0.1)	Aubergines (6)	Beans, French (39)	Brussels sprouts (147)
Beef, steak (0.8)	Baked beans (3)	Beans, broad (19)	Cabbage, green (339)
Bread, white (0.4)	Barley (7)	Cabbage, red (19)	Kale (618)
Chicken, thigh (0.1)	Beef, corned (7)	Cauliflower (13)	Lettuce (129)
Coconut oil (0.5)	Beef, minced (2)	Cockpeas (21)	Parsley (548)
Cod, fresh, fillet (0.1)	Bilberries (4)	Cucumber (21)	Rapeseed oil (123)
Corn flakes (<0.1)	Bran, wheat (10)	Greengages (15)	Soybean oil (173)
Flour, white (0.8)	Bread, whole meal (2)	Mustard greens, cress (88)	Spinach (380)
Grapefruit (<0.1)	Butter (7)	Olive oil, extra virgin (80)	Watercress (315)
Ham, tinned (0.1)	Carrots (6)	Peas (34)	
Maize (0.3)	Cheeses, various (2–6)		
Mangoes (0.5)	Chocolate, plain (2)		
Melon, yellow (0.1)	Corn oil (3)		
Melon, water (0.3)	Courgettes (3)		
Milk, cows (0.6)	Cranberries (2)		
Mushrooms (0.3)	Cream, double (6)		
Oranges (<0.1)	Dates, fresh (6)		
Parsnips (<0.1)	Egg yolk (2)		
Peanuts, roast (0.4)	Figs, fresh (3)		
Pilchards, brine (0.6)	Grapes, black (8)		
Pineapple (0.2)	Grapes, green (9)		
Pork, chop, lean (<0.1)	Leeks (10)		
Potatoes (0.9)	Liver, lamb (7)		
Rice, white (0.1)	Liver, ox (4)		
Rice, brown (0.8)	Nectarines (3)		
Salmon, tin, brine (0.1)	Oats (10)		
Sausage, pork/beef (0.2)	Palm oil (8)		
Spaghetti (0.2)	Peaches, fresh (4)		
Tuna, tin, brine (0.1)	Pears (6)		
Turnips (0.2)	Peppers, green (6)		
Yoghurt (0.8)	Peppers, red (2)		
	Plums, red (8)		
	Raisins (4)		
	Rhubarb (4)		
	Safflower oil (3)		
	Sunflower oil (6)		
	Swede (2)		
	Tomatoes (6)		
	Wheat (8)		

Note: Data are for raw foods except where cooked form is indicated. Numbers in parentheses indicate the micrograms of phylloquinone per 100 g (3.5 oz.) food.

Source: Adapted from Shearer, Bach, and Kohlmeier (1996: 1183S). Reprinted with permission.

Table IV.A.6.2 represent analyses of a single sample. Phylloquinone concentrations in plants, however, are affected by the stage of maturation and the geographic location of the plant. Further, different parts of the plant may differ in phylloquinone content. Outer leaves of the cabbage, for example, contain three to six times more phylloquinone than inner leaves (Shearer et al. 1980; Ferland and Sadowski 1992a). Although tea leaves and regular ground coffee contain high concentrations of phylloquinone, the brews are not a dietary source of the vitamin (Booth, Madabushi, et al. 1995). The phylloquinone content of oils declines slightly with heat and rapidly on exposure to daylight and fluorescent light when the oils are stored in transparent containers (Ferland and Sadowski 1992b; Shearer et al. 1996).

Recent evidence indicates that the intestinal absorption of phylloquinone from vegetables is poor and is improved by the simultaneous ingestion of fat (Vermeer et al. 1996). Bile acids, secreted into the intestines in response to fat in a meal, are necessary for the absorption of the fat-soluble K vitamins. There is a need for additional studies on the bioavailability of the K vitamins in foods to complement our knowledge of their vitamin K content (Vermeer et al. 1996).

Less is known about the vitamin K content of menaquinones (synthesized by bacteria) in foods. Livers of ruminant species, such as cows, have been found to have nutritionally significant quantities of menaquinones, but the concentrations of menaquinones in other animal organs were very low (Hirauchi et al. 1989; Shearer et al. 1996). Cheeses contained significant quantities of two different menaquinones, but very low quantities of menaquinones were found in milk and yoghurt (Shearer et al. 1996).

A topic currently in dispute is whether the menaquinones produced by bacterial action in the gut are substantially utilized by humans. Concentrations of various menaquinones have been found to be higher than that of phylloquinone in normal human livers (Usui et al. 1989) and, importantly, antibiotics have been known since the late 1940s and early 1950s to sometimes create a vitamin K deficiency. Nonetheless, the degree of importance of menaquinones in human nutrition has not yet been determined (Suttie 1995), and there are data to suggest that the long-chain menaquinones found in human liver are not as effective a form of the vitamin as phylloquinone. Moreover, elimination of foods high in vitamin K from a normal diet can produce signs of vitamin K insufficiency, suggesting that bacterial synthesis of menaquinones only partially satisfies the human vitamin K requirement (Suttie 1995).

Documented cases of vitamin K deficiency in adults have been uncommon. The National Research Council of the National Academy of Sciences recommends a daily dietary allowance (RDA) of approximately 1 microgram (μg) vitamin K per kilogram (kg) of body weight (National Research Council 1989).

Thus, the requirement for an individual weighing 150 pounds (68 kg) would be 68 μg phylloquinone. The average vitamin K intake of 10 male college students was determined to be 77 $\mu\text{g}/\text{day}$ based on analysis of food composites (Suttie et al. 1988) – an average intake that corresponds to 1.04 μg vitamin K/kg body weight. Foods rich in vitamin K, such as spinach, broccoli, and Brussels sprouts, were not usually consumed in significant amounts by these students. This suggests that daily vitamin K intake may vary considerably depending on whether foods rich in vitamin K are consumed (Suttie et al. 1988). Table IV.A.6.2 makes it clear that 100 grams (g) of spinach (380 μg phylloquinone per 100 g) would contribute much to the recommended daily dietary intake of vitamin K, even if phylloquinone is poorly absorbed from this food.

Vitamin K deficiency can result from inadequate intestinal absorption. Because vitamin K is a fat-soluble vitamin, bile acids and pancreatic juice are needed for its absorption. Thus, causes of poor intestinal absorption can be insufficient production of bile acids, inadequate release of bile acids (for example, obstruction of the bile ducts by gallstones), or pancreatic insufficiency (Report of the Committee on Nutrition 1961; Suttie 1991). Levels of all the K vitamins were reduced in patients with chronic hepatitis and liver cirrhosis (Usui et al. 1989). In addition, secondary vitamin K deficiency has been observed in human subjects taking megadoses of vitamin E (Korsan-Bengtson, Elmfeldt, and Holm 1974). Vitamin E supplementation is currently on the rise because of its role as an antioxidant. However, the effect of large doses of vitamin E on vitamin K status must always be borne in mind.

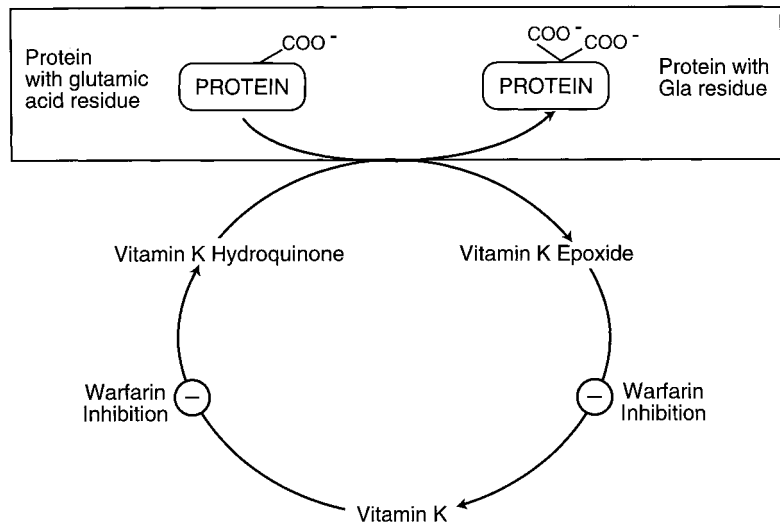
An upper limit for vitamin K intake has not been set, but 20 $\mu\text{g}/100$ kilocalories has been suggested by J. W. Suttie and R. E. Olson (Olson 1989). Menadiolone, which is not a natural form of vitamin K, can cause severe toxic reactions in infants if administered in large doses. Phylloquinone, however, has been given without adverse effects to infants in a single intramuscular dose that is 100 times their RDA. Adult dietary intakes of 10 to 15 times the RDA also cause no adverse effects (Olson 1989).

Function of Vitamin K

Vitamin K is involved in the synthesis of four clotting factors, one procoagulant protein, two anticoagulant proteins, and two bone proteins. The vitamin K-dependent clotting factors are integral to blood coagulation, such that an untreated vitamin K deficiency results in death due to bleeding. There also must be control of coagulation, and the body, therefore, has anticoagulant systems. Two of the known anticoagulant proteins are vitamin K-dependent.

The vitamin K-dependent bone proteins (in contrast to the vitamin K-dependent clotting proteins) have only recently been discovered, and the physiological role of these proteins in bone has not yet been determined.

Figure IV.A.6.1. The vitamin K-dependent carboxylase reaction. The vitamin K-dependent carboxylase converts glutamic acid residues of specific proteins to gamma-carboxyglutamic acid (Gla) residues. Vitamin K hydroquinone is simultaneously converted to vitamin K epoxide. The epoxide is reconverted to vitamin K, which is converted to vitamin K hydroquinone, and the cycle repeats. Warfarin inhibits conversion of the epoxide to vitamin K and conversion of vitamin K to vitamin K hydroquinone.



Mechanism of Action of Vitamin K

Vitamin K is a cofactor for an enzyme (vitamin K-dependent carboxylase) that carboxylates specific glutamic acid residues of proteins, converting these amino acid residues to gamma-carboxyglutamic acid residues (Figure IV.A.6.1). This modification occurs after the protein has been synthesized. The gamma-carboxyglutamyl residues are abbreviated Gla.

Although it was established in the early 1950s that vitamin K was necessary for the synthesis of the clotting proteins prothrombin (factor II) and factors VII, IX, and X, the vitamin K-dependent carboxylation reaction was not discovered until 1974. Investigators at that time demonstrated the presence of a new amino acid, gamma-carboxyglutamic acid, in prothrombin (Nelsentuen, Zytkevich, and Howard 1974; Stenflo et al. 1974). The nature of the carboxylase reaction was soon delineated (Shah and Suttie 1974; Esmon, Sadowski, and Suttie 1975; Suttie 1985). A reduced form of vitamin K (vitamin K hydroquinone) is converted to vitamin K epoxide as the glutamic acid residues of the specific protein are converted to gamma-carboxyglutamic acid residues (Figure IV.A.6.1). At physiological concentrations, vitamin K must be regenerated and reutilized and is recycled. Two reactions take place sequentially to convert vitamin K epoxide to its reduced form (vitamin K hydroquinone) for continued carboxylation (Figure IV.A.6.1). These two reactions are prevented by coumarin anticoagulants, synthetic compounds such as warfarin, which are used medically to counteract excessive clotting.

The vitamin K-dependent (Gla-containing) proteins that have been well characterized include prothrombin and the clotting factors VII, IX, and X (plasma proteins synthesized by the liver); protein Z, a procoagulant plasma protein; protein C and protein S,

plasma anticoagulants synthesized by the liver; osteocalcin, also called bone Gla protein; and matrix Gla protein, a bone-matrix protein.

Vitamin K-Dependent Proteins

The clotting proteins. When a cut or injury occurs, platelets (cells in blood) converge on the injury to form a plug, and a clot forms on the platelet plug. This action stops blood loss and prevents the injured person from bleeding to death. A clot is formed by a series of transformations involving more than ten different proteins. In the final stage, fibrinogen, a soluble blood protein, is cleaved by thrombin and converted to fibrin. Fibrin monomers are then cross-linked to form the insoluble or hard clot. A schematic of the clotting cascade is illustrated in Figure IV.A.6.2. Hundreds of papers have been published on the blood clotting cascade; reviews by C. M. Jackson and Y. Nemerson (1980) and by B. Furie and B. C. Furie (1992) provide additional information.

The proteins of the clotting cascade that are vitamin K-dependent are prothrombin (factor II), factor VII, factor IX, and factor X (Figure IV.A.6.2). These proteins, which are central to the blood clotting cascade, are carboxylated after synthesis in a reaction requiring vitamin K (Uotila 1990). The Gla residues make the proteins more negative and endow them with an increased ability to bind positively charged calcium ions. Calcium ions serve as bridges between these proteins and the negatively charged phospholipids of the platelet membrane. The proteins are brought in close proximity to each other on the platelet membrane, augmenting their activation. In the absence of vitamin K, carboxylation of prothrombin and factors VII, IX, and X does not occur. Activation of these factors proceeds so slowly that bleeding may result.

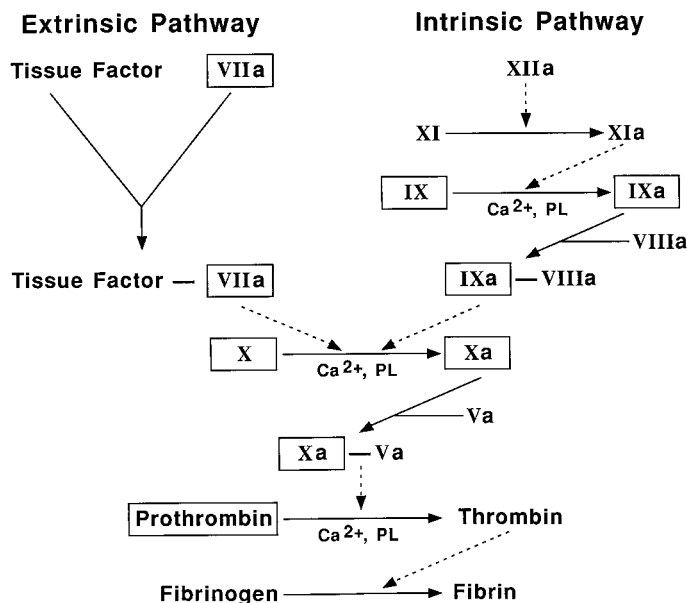


Figure IV.A.6.2. An outline of the clotting sequence. Dashed lines indicate an enzyme action, and solid lines indicate the activation of a protein, formation of a complex, or conversion of one compound to another. The presence of an "a" indicates that a factor has been activated. Vitamin K-dependent factors are located in rectangles. Calcium ions (Ca^{2+}) and phospholipids (PL), supplied by the platelet membrane, are required in those reactions that involve the vitamin K-dependent cofactors.

In response to a vessel injury, platelets aggregate at the injured site and form a platelet plug. The platelet membrane provides a surface on which a clot will subsequently form. Tissue factor (a protein) is also released at the injured site. It combines with activated factor VII. This complex (tissue factor-factor VIIa) activates factor X, which combines with activated factor V. The factor Xa-factor Va complex is able to activate prothrombin (factor II) to produce thrombin, an active enzyme that cleaves fibrinogen molecules to produce fibrin monomers. The fibrin monomers are cross-linked to produce the hard clot. The tissue factor-factor VIIa pathway for activating factor X is called the extrinsic pathway.

Factor X is also activated by a pathway called the intrinsic pathway. All proteins of the pathway are located in plasma. Factor XII, which is perhaps activated by collagen fibers from the injured blood vessel, activates factor XI. Factor XIa activates factor IX, which then combines with activated factor VIII to produce a factor IXa-factor VIIIa complex, which activates factor X.

At the time of identification of vitamin K as the antihemorrhagic factor, prothrombin and fibrinogen were the only proteins in the clotting cascade characterized as involved in the formation of the fibrin clot. The laboratories of Dam (Dam, Schönheyder, and Tage-Hansen 1936; Dam 1964) and A. J. Quick (Quick 1937) independently demonstrated that the activity of prothrombin was decreased in the plasma of chicks fed hemorrhagic diets. The choice of the chick as the experimental animal in these early studies was fortuitous, since chicks develop the symptoms of vitamin K deficiency more readily than other experimental animals (Suttie 1991). Factors VII, IX, and X, the other vitamin K-dependent clotting factors, were not established as essential plasma proteins and vitamin K-dependent proteins until the 1950s.

Human protein Z, another vitamin K-dependent protein, was purified and first described in 1984 (Broze and Miletich 1984). It is synthesized by the liver and promotes the association of thrombin (activated prothrombin) with phospholipid surfaces, a process necessary for clotting. Protein Z deficiency has recently been described as a new type of bleeding tendency (Kemkes-Matthes and Matthes 1995).

The anticoagulants. For several decades prior to the discovery of protein C (Stenflo 1976), the only known vitamin K-dependent plasma proteins were the clotting proteins II, VII, IX, and X. Protein S was subsequently isolated and purified (Di Scipio et al. 1977). Protein C and protein S are now known to be components of a very important anticoagulant system in plasma (Figure IV.A.6.3). Protein C, when activated and bound to protein S, is able to cleave, and thereby inactivate, two of the activated clotting factors, factors Va and VIIIa (Walker 1980; Stenflo 1984; Esmon 1987). The presence of both coagulant and anticoagulant systems in plasma allows for control of clotting.

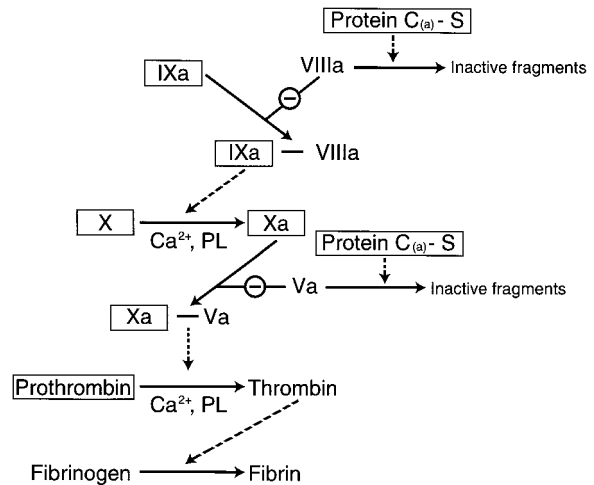
Hereditary protein C deficiency and protein S deficiency have been discovered (Griffin et al. 1981; Pabinger 1986; Miletich, Sherman, and Broze 1987; Broekmans and Conrad 1988; Bertina 1989; Preissner 1990; Rick 1990). The clinical manifestation of protein S or protein C deficiency is thrombosis (excessive clotting).

The bone proteins. Until the discovery of osteocalcin (bone Gla protein, BGP) in the 1970s, vitamin K was assumed to function only in coagulation. The Gla residues of osteocalcin, a major protein of bone, pro-

Figure IV.A.6.3. The vitamin K-dependent anticoagulant system.

Dashed lines indicate an enzyme action, and solid lines indicate the activation of a protein, formation of a complex, or conversion of one compound to another. The presence of an "a" indicates that a factor has been activated. Vitamin K-dependent factors are located in rectangles.

A complex of activated protein C and protein S cleaves activated factor VIII and activated factor V, rendering them inactive. The inactivation of these two factors causes the cessation of blood coagulation, due to decreases in the rates of the two reactions with the encircled negative signs.



vide a point of interaction between the protein and bone mineral (Hauschka, Lian, and Gallop 1975; Poser and Price 1979). Matrix Gla protein was isolated and identified in 1983 (Price, Urist, and Otawara 1983). The synthesis of osteocalcin and matrix Gla protein is regulated by the hormonal form of vitamin D (1,25-dihydroxyvitamin D) (Price 1988). Two vitamins, vitamins K and D, are thus involved in the synthesis of osteocalcin and matrix Gla protein.

A fetal warfarin syndrome, first reported in 1975, has been found in infants born to mothers on warfarin anticoagulant therapy during pregnancy (Becker et al. 1975; Pettifor and Benson 1975; Shaul, Emery, and Hall 1975). Some of the abnormalities identified have been associated with bone formation. Although the functions of osteocalcin and matrix Gla protein are not yet established, defective carboxylation of these two proteins in the presence of warfarin may be a cause of the bone abnormalities associated with fetal warfarin syndrome (Price et al. 1983; Pauli et al. 1987; Price 1988).

In recent years, several studies have implicated vitamin K insufficiency in the pathogenesis of osteoporosis (Binkley and Suttie 1995; Vermeer, Jie, and Knapen 1995; Shearer et al. 1996; Vermeer et al. 1996). Osteoporosis is an age-related disorder characterized by inadequate skeletal strength due to bone loss (decreased bone density) and a predisposition to bone fractures. Peak bone mass is achieved between ages 25 and 35, and bone loss is a natural process that begins thereafter. The highest rate of bone loss for women, however, occurs within the first three to six years after the onset of menopause. Women are at greater risk for osteoporosis than men because of a lower peak bone density and the rapid postmenopausal loss of bone, related, in part, to reduced estrogen levels.

Low concentrations of vitamin K have been found in the blood of patients with bone fractures. When the vitamin K status is insufficient, vitamin K-dependent proteins tend to be undercarboxylated. Osteocalcin is

more sensitive to a low vitamin K intake than the coagulation proteins, and, thus, undercarboxylated osteocalcin is the most sensitive known marker of vitamin K insufficiency. Undercarboxylated osteocalcin has been reported to be increased in postmenopausal women and to be highest in those individuals with the lowest hip bone density and greatest hip fracture risk. Vitamin K supplementation has been shown to decrease bone loss, and clinical trials have been initiated to test the effect of long-term vitamin K supplementation on bone mass. Measurements of undercarboxylated osteocalcin suggest that vitamin K insufficiency, formerly characterized by the status of the coagulation system, may be more common than previously thought (Binkley and Suttie 1995; Vermeer et al. 1996).

Other proteins. Gla-containing proteins have also been purified from kidney (Griep and Friedman 1980), urine (Nakagawa et al. 1983), liver mitochondria (Gardemann and Domagk 1983), and human spermatozoa (Soute et al. 1985). The function of these proteins is not presently known.

Abnormal Carboxylation

Hereditary deficiency of vitamin K-dependent clotting proteins (factors II, VII, IX, and X) can result in a rare bleeding disorder that has so far been diagnosed in only seven patients. The first reports of this disorder were in 1966 (Fischer and Zweymuller 1966; McMillan and Roberts 1966), and data (Brenner et al. 1990) suggest that the defect is abnormal carboxylation (due to a defective liver vitamin K-dependent carboxylase enzyme). Deficiencies in the plasma anticoagulants protein C and protein S have also been observed (Brenner et al. 1990).

Hemorrhagic Diseases of Infants

Although vitamin K deficiency symptoms are not very prevalent in adults, hemorrhagic diseases related to

vitamin K deficiency are observed in infants worldwide. Three patterns of vitamin K-deficiency hemorrhage – early hemorrhagic disease of the newborn (HDN), classic hemorrhagic disease of the newborn, and late hemorrhagic disease – have been described. The frequency of these diseases, particularly that of late hemorrhagic disease, has increased during the last decade (Lane and Hathaway 1985; Kries, Shearer, and Göbel 1988; Greer 1995).

Early Hemorrhagic Disease

Early hemorrhagic disease of the newborn is characterized by severe, and sometimes life-threatening, hemorrhage at the time of delivery or during the first 24 hours after birth (Lane and Hathaway 1985; Kries et al. 1988). The bleeding varies from skin bruising, cephalohematoma, or umbilical bleeding to widespread and fatal intracranial, intra-abdominal, intrathoracic, and gastrointestinal bleeding (Lane and Hathaway 1985). The disease is often seen in infants whose mothers have taken drugs during pregnancy that affect vitamin K metabolism, such as therapeutic doses of warfarin. Idiopathic cases (no known cause) have also been reported.

Classic HDN

C. W. Townsend (1894) was the first person to use the term “haemorrhagic disease of the newborn.” He described 50 infants who began bleeding on the second or third day of life, most commonly from the gastrointestinal tract, and speculated that the disease was of infectious origin. Dam and colleagues (1952) studied 33,000 infants and concluded that low levels of the clotting factor prothrombin in newborns was secondary to a vitamin K deficiency and occurred primarily in breast-fed infants. The researchers showed that this problem could be prevented by administration of vitamin K to mothers prior to delivery and to the infant shortly after delivery.

Normal full-term infants have reduced blood concentrations of the vitamin K-dependent clotting factors II, VII, IX, and X. Further, the levels of these factors decline during the first few days of life. Hemorrhagic disease of the newborn, often called early vitamin K deficiency, occurs from 1 to 7 days after birth. Breast-fed infants are at risk because of the low vitamin K content of breast milk. Bleeding occurs in the gastrointestinal tract, skin, and nose and at circumcision. The disease is prevented by vitamin K prophylaxis at birth (Lane and Hathaway 1985; Kries et al. 1988; Greer 1995), as indicated in 1952 (Dam et al. 1952).

Late Hemorrhagic Disease

Late hemorrhagic disease (late neonatal vitamin K deficiency) strikes at some infants (1 to 12 months of age) who are predominantly breast-fed and who do not receive vitamin K supplementation (Lane and Hathaway 1985; Hanawa et al. 1988; Kries et al. 1988).

It is characterized by intracranial, skin, and gastrointestinal bleeding.

That late hemorrhagic disease of the neonate has been observed primarily in breast-fed infants may result from the low vitamin K content of human milk (Lane and Hathaway 1985; Kries et al. 1988; Canfield et al. 1990). However, even if administered at birth, vitamin K may not always prevent deficiency in older infants because of their metabolism rates and consequent elimination of the vitamin. In Japan, from January 1981 to June 1985, 543 cases of vitamin K deficiency were reported in infants over 2 weeks of age (Hanawa et al. 1988). Of these, 427 were diagnosed as having idiopathic vitamin K deficiency, and 387 (90 percent) of this group had been entirely breast-fed (Hanawa et al. 1988). The concentrations of phyloquinone in human milk (mean 2.1 µg/liter) have been found to be significantly lower than those found in cows' milk (mean 4.9 µg/liter) and in unsupplemented infant formulas containing only fat from cows' milk (mean 4.2 µg/liter). Supplemented infant formulas, by contrast, had higher levels of phyloquinone than cows' milk; two of those studied had levels of 75.1 µg/liter and 101.8 µg/liter (Harroon et al. 1982; Kries et al. 1987).

It has been proposed that the late neonatal vitamin K deficiency that occurs in predominantly breast-fed infants may not result solely from lower vitamin K intake but from its combination with other factors, including subclinical liver dysfunction (Hanawa et al. 1988; Kries and Göbel 1988; Matsuda et al. 1989). In fact, I. Matsuda and colleagues (1989) have suggested that the higher content of vitamin K in formulas, unlike the lower content of vitamin K in human milk, can actually mask such defects.

In another study, phyloquinone levels and coagulation factors were measured in healthy term newborns until 4 weeks after birth (Pietersma-de Bruyn et al. 1990). Although breast-fed infants had lower serum phyloquinone levels than formula-fed infants, the levels of the vitamin K-dependent clotting factors II and X were comparable in the two groups. Thus, the authors of this study concluded that breast milk contains sufficient vitamin K for optimal carboxylation of the clotting factors. They have also suggested that the vitamin K supplementation of infant formulas results in higher than normal serum levels of vitamin K in infants without a concomitant rise in the levels of the vitamin K-dependent clotting proteins (Pietersma-de Bruyn et al. 1990).

Vitamin K Prophylaxis

Prophylactic use of vitamin K has been recommended for all newborns by the American Academy of Pediatrics since 1961. Vitamin K prophylaxis is standard for all infants in some geographical regions but is, or has been, “selective” (given only to at-risk infants) in others. However, in some regions where the “selective policy” was in effect, there was a resur-

gence (in the 1980s and 1990s) of hemorrhagic disease (Tulchinsky et al. 1993; Greer 1995).

The mode of administration (oral versus intramuscular) of vitamin K also became an issue of controversy in the 1990s because of reports from one research group (unconfirmed by two other groups) suggesting an association between intramuscular administration of vitamin K to the neonate and the subsequent development of childhood cancers. But in several countries (for example, Sweden, Australia, and Germany) where use of oral vitamin K prophylaxis (compared with intramuscular) has increased, there has been an increased incidence of late hemorrhagic disease. Oral prophylaxis prevents classic hemorrhagic disease but is less effective in preventing late hemorrhagic disease. Prevention of the latter requires repeated oral doses of vitamin K during the first 2 months of life for exclusively breast-fed infants or other infants at risk (for example, infants with liver disease), all of which presents problems with compliance (Greer 1995; Thorp et al. 1995).

The Vitamin K Ad Hoc Task Force of the American Academy of Pediatrics (1993) has recommended that to prevent hemorrhagic disease, phyloquinone be administered at birth to all newborns as a single, intramuscular dose of 0.5 to 1 milligram. The Task Force has also recommended that research be done to judge the efficacy, safety, and bioavailability of oral formulations of vitamin K. Oral supplements of vitamin K to nursing mothers might become an alternative method of prophylaxis to prevent late hemorrhagic disease (Greer 1995). Physicians could also advise pregnant and nursing mothers to increase their intake of green, leafy vegetables.

The Oral Anticoagulants

The discovery of the first coumarin anticoagulant, dicoumarol, has been interestingly described by K. P. Link (1959). During a blizzard on a Saturday afternoon in February 1933, a Wisconsin farmer appeared at the University of Wisconsin Biochemistry Building with a dead cow, a milk can containing blood without clotting capacity, and about 100 pounds of spoiled sweet clover hay. The farmer's cows were suffering from "sweet clover disease," a malady caused by feeding cattle improperly cured hay made from sweet clover. If the type of hay was not changed and if the cows were not transfused, they developed a prothrombin deficit and bled to death. Link and his colleagues began to work on the isolation and identification of the hemorrhagic agent, and in 1939, it was identified as dicoumarol, a coumarin derivative. As a result of the spoilage, coumarin, a natural component of sweet clover, had been converted to dicoumarol (Link 1959).

Since the discovery of dicoumarol, other coumarin derivatives have been synthesized (Link 1959). The most famous of the coumarin derivatives is warfarin,

named for the University of Wisconsin Alumni Research Foundation (WARF), which received a patent for the compound. The coumarin compounds interfere with the metabolism of vitamin K, preventing conversion of vitamin K epoxide to vitamin K and also preventing reduction of vitamin K to vitamin K hydroquinone (Figure IV.A.6.1). This interference causes a buildup of vitamin K epoxide, and when physiological levels of vitamin K are present, the inability of the epoxide to be converted to the vitamin creates a relative vitamin K deficiency. The vitamin K-dependent proteins, including the clotting proteins II, VII, IX, and X, are, therefore, not carboxylated – or only partially carboxylated – and remain inactive (Suttie 1990).

The coumarin compounds, particularly warfarin, have been used as oral anticoagulants in the long-term treatment of patients prone to thrombosis (clot formation). The formation of a clot in a coronary artery narrowed by atherosclerosis is a causative factor in the development of acute myocardial infarction (heart attack), and warfarin is used after acute myocardial infarction to prevent further thrombus (clot) formation and reinfarction.

A recent study of the effect of warfarin on mortality and reinfarction after myocardial infarction concluded that "long-term therapy with warfarin has an important beneficial effect after myocardial infarction" (Smith, Arnesen, and Holme 1990: 147). This study involved 607 patients treated with warfarin and 607 patients treated with a placebo. Compared with the placebo group, warfarin therapy reduced the number of deaths by 24 percent, the number of reinfarctions by 34 percent, and the number of cerebrovascular accidents (strokes) by 55 percent.

In anticoagulant therapy, a dose of the anticoagulant is selected that achieves effective anticoagulation and minimizes bleeding complications (Second Report of the Sixty Plus Reinfarction Study Research Group 1982). The advantage of using warfarin for long-term therapy is that an overdosage is corrected by the administration of a large dose of vitamin K. Large doses of vitamin K are converted to vitamin K hydroquinone by enzymes that are not inhibited by warfarin (Link 1959; Suttie 1990). Correspondingly, persons undergoing warfarin anticoagulant therapy should avoid a diet that is high in vitamin K by limiting, among other things, the intake of green, leafy vegetables (Pedersen et al. 1991).

A Note on Hemophilia

Hemophilia is a bleeding disorder caused by a deficiency in one or, more rarely, two of the clotting proteins (De Angelis et al. 1990). The majority (85 percent) of hemophilia patients are deficient in factor VIII (hemophilia A), with 10 to 12 percent deficient in factor IX (hemophilia B). Rarer forms of the disease result from deficiency in other clotting factors. The

hemophilia patient must receive injections of the missing protein in order to maintain proper coagulation. Hemophilia is a genetic disease, most commonly transmitted as a sex-linked recessive trait, and is not related to vitamin K status (Furie and Furie 1988).

Myrtle Thierry-Palmer

Bibliography

- Almquist, H. J. 1979. Vitamin K: Discovery, identification, synthesis, functions. *Federation Proceedings* 38: 2687-9.
- Almquist, H. J., and E. L. R. Stokstad. 1935. Dietary haemorrhagic disease in chicks. *Nature* 136: 31.
- Becker, M. H., N. B. Genieser, M. Finegold, et al. 1975. Chondrodysplasia punctata: Is maternal warfarin therapy a factor? *American Journal of Diseases of Children* 129: 356-9.
- Bertina, R. M. 1989. The control of hemostasis: Protein C and protein S. *Biomedical Progress* 4: 53-6.
- Binkley, N. C., and J. W. Suttie. 1995. Vitamin K nutrition and osteoporosis. *Journal of Nutrition* 125: 1812-21.
- Booth, S. L., H. T. Madabushi, K. W. Davidson, and J. A. Sadowski. 1995. Tea and coffee brews are not dietary sources of vitamin K-1 (phylloquinone). *Journal of the American Dietetic Association* 95: 82-3.
- Booth, S. L., J. A. Sadowski, and J. A. T. Pennington. 1995. Phylloquinone (vitamin K₁) content of foods in the U.S. Food and Drug Administration's Total Diet Study. *Journal of Agricultural and Food Chemistry* 43: 1574-9.
- Booth, S. L., J. A. Sadowski, J. L. Weihrauch, and G. Ferland. 1993. Vitamin K₁ (phylloquinone) content of foods: A provisional table. *Journal of Food Composition and Analysis* 6: 109-20.
- Brenner, B., S. Tavori, A. Zivelin, et al. 1990. Hereditary deficiency of all vitamin K-dependent procoagulants and anticoagulants. *British Journal of Haematology* 75: 537-42.
- Broekmans, A. W., and J. Conrad. 1988. Hereditary protein C deficiency. In *Protein C and related proteins*, ed. R. M. Bertina, 160-81. New York.
- Broze, G. J., Jr., and J. P. Miletich. 1984. Human protein Z. *Journal of Clinical Investigation* 73: 933-8.
- Canfield, L. M., J. M. Hopkinson, A. F. Lima, et al. 1990. Quantitation of vitamin K in human milk. *Lipids* 25: 406-11.
- Dam, H. 1935. The antihemorrhagic vitamin of the chick. Occurrence and chemical nature. *Nature* 135: 652-3.
1942. Vitamin K, its chemistry and physiology. *Advances in Enzymology* 2: 285-324.
1964. The discovery of vitamin K, its biological functions and therapeutic application. In *Nobel lectures: Physiology and medicine 1942-1962*, ed. The Nobel Foundation, 8-26. Amsterdam.
- Dam, H., H. Dyggve, H. Larsen, and P. Plum. 1952. The relation of vitamin K deficiency to hemorrhagic disease of the newborn. *Advances in Pediatrics* 5: 129-53.
- Dam, H., F. Schönheyder, and E. Tage-Hansen. 1936. Studies on the mode of action of vitamin K. *Biochemical Journal* 30: 1075-9.
- De Angelis, V., B. M. Orazi, L. Santarossa, and G. Molaro. 1990. Combined factor VIII and factor XI congenital deficiency: A case report. *Haematologica* 75: 272-3.
- Di Scipio, R. G., M. A. Hermanson, S. G. Yales, and E. W. Davie. 1977. A comparison of human prothrombin, factor IX (Christmas factor), factor X (Stuart factor), and protein S. *Biochemistry* 16: 698-706.
- Doisy, E. A., S. B. Binkley, and S. A. Thayer. 1941. Vitamin K. *Chemical Reviews* 28: 477-517.
- Esmon, C. T. 1987. The regulation of natural anticoagulant pathways. *Science* 235: 1348-52.
- Esmon, C. T., J. A. Sadowski, and J. W. Suttie. 1975. A new carboxylation reaction. The vitamin K-dependent incorporation of H¹⁴CO₂ into prothrombin. *Journal of Biological Chemistry* 250: 4744-8.
- Ferland, G., and J. A. Sadowski. 1992a. Vitamin K₁ (phylloquinone) content of edible oils: Effects of heating and light exposure. *Journal of Agricultural and Food Chemistry* 40: 1869-73.
- 1992b. Vitamin K₁ (phylloquinone) content of green vegetables: Effects of plant maturation and geographical growth location. *Journal of Agricultural and Food Chemistry* 40: 1874-7.
- Fischer, M., and E. Zweymüller. 1966. Kongenitaler kombinierter Mangel der Faktoren II, VII, und X. *Zeitschrift für Kinderheilkunde* 95: 309-23.
- Furie, B., and B. C. Furie. 1988. The molecular basis of blood coagulation. *Cell* 53: 505-18.
1992. Molecular and cellular biology of blood coagulation. *The New England Journal of Medicine* 326: 800-6.
- Gardemann, A., and G. F. Domagk. 1983. The occurrence of Γ -carboxyglutamate in a protein isolated from ox liver mitochondria. *Archives of Biochemistry and Biophysics* 220: 347-53.
- Greer, F. R. 1995. Vitamin K deficiency and hemorrhage in infancy. *Perinatal Hematology* 22: 759-77.
- Griep, A. E., and P. A. Friedman. 1980. Purification of a protein containing Γ -carboxyglutamic acid from bovine kidney. In *Vitamin K metabolism and vitamin K-dependent proteins*, ed. J. W. Suttie, 307-10. Baltimore, Md.
- Griffin, J. H., B. Evatt, T. S. Zimmerman, et al. 1981. Deficiency of protein C in congenital thrombotic disease. *Journal of Clinical Investigations* 68: 1370-3.
- Hanawa, Y., M. Maki, B. Murata, et al. 1988. The second nationwide survey in Japan of vitamin K deficiency in infancy. *European Journal of Pediatrics* 147: 472-7.
- Harroon, Y., M. J. Shearer, S. Rahim, et al. 1982. The content of phylloquinone (vitamin K₁) in human milk, cows' milk and infant formula determined by high-performance liquid chromatography. *Journal of Nutrition* 112: 1105-17.
- Hauschka, P. V., J. B. Lian, and P. M. Gallop. 1975. Direct identification of the calcium-binding amino acid Γ -carboxyglutamate in mineralized tissue. *Proceedings of the National Academy of Science* 72: 3925-9.
- Hirauchi, K., T. Sakano, S. Notsumoto, et al. 1989. Measurement of K vitamins in animal tissues by high-performance liquid chromatography with fluorimetric detection. *Journal of Chromatography* 497: 131-7.
- Holst, W. F., and E. R. Halbrook. 1933. A "scurvy-like" disease in chicks. *Science* 77: 354.
- Jackson, C. M., and Y. Nemerson. 1980. Blood coagulation. *Annual Review of Biochemistry* 49: 765-811.
- Kemkes-Matthes, B., and K. J. Matthes. 1995. Protein Z deficiency: A new cause of bleeding tendency. *Thrombosis* 79: 49-55.
- Korsan-Bengtson, K., D. Elmfeldt, and T. Holm. 1974. Prolonged plasma clotting time and decreased fibrinolysis after long-term treatment with α -tocopherol. *Thrombosis Diathesis Haemorrhagia* 31: 505-12.
- Kries, R. von, and U. Göbel. 1988. Vitamin K prophylaxis:

- Oral or parenteral? (Letter). *American Journal of Diseases of Children* 142: 14-15.
- Kries, R. von, M. J. Shearer, and U. Göbel. 1988. Vitamin K in infancy. *European Journal of Pediatrics* 147: 106-12.
- Kries, R. von, M. J. Shearer, P. T. McCarthy, et al. 1987. Vitamin K content of maternal milk: Influence of the stage of lactation, lipid composition, and vitamin K₁ supplements given to the mother. *Pediatric Research* 22: 513-17.
- Lane, P. A., and W. E. Hathaway. 1985. Vitamin K in infancy. *Journal of Pediatrics* 106: 351-9.
- Link, K. P. 1959. The discovery of dicoumarol and its sequels. *Circulation* 19: 97-107.
- Matsuda, I., S. Nishiyama, K. Motohara, et al. 1989. Late neonatal vitamin K deficiency associated with subclinical liver dysfunction in human milk-fed infants. *Journal of Pediatrics* 114: 602-5.
- McFarlane, W. D., W. R. Graham, Jr., and G. E. Hall. 1931. Studies in protein nutrition of the chick. I. The influence of different protein concentrates on the growth of baby chicks, when fed as the source of protein in various simplified diets. *Journal of Nutrition* 4: 331-49.
- McMillan, C. W., and H. R. Roberts. 1966. Congenital combined deficiency of coagulation factors II, VII, IX and X. *New England Journal of Medicine* 274: 1313-15.
- Miletich, J., L. Sherman, and G. Broze, Jr. 1987. Absence of thrombosis in subjects with heterozygous protein C deficiency. *New England Journal of Medicine* 317: 991-6.
- Nakagawa, Y., V. Abram, F. J. Kezdy, et al. 1983. Purification and characterization of the principal inhibitor of calcium oxalate monohydrate crystal growth in human urine. *Journal of Biological Chemistry* 258: 12594-600.
- National Research Council (Subcommittee on the Tenth Edition of the RDAs). 1989. *Recommended dietary allowances*. Tenth edition. Washington, D.C.
- Nelsestuen, G. L., T. H. Zytkevich, and J. B. Howard. 1974. The mode of action of vitamin K. Identification of gamma-carboxyglutamic acid as a component of prothrombin. *Journal of Biological Chemistry* 249: 6347-50.
- Olson, J. A. 1989. Upper limits of vitamin A in infant formulas, with some comments on vitamin K. *Journal of Nutrition* 119: 1820-4.
- Pabinger, I. 1986. Clinical relevance of protein C. *Blut* 53: 63-75.
- Pauli, R. M., J. B. Lian, D. F. Mosher, and J. W. Suttie. 1987. Association of congenital deficiency of multiple vitamin K-dependent coagulation factors and the phenotype of the warfarin embryopathy: Clues to the mechanism of teratogenicity of coumarin derivatives. *American Journal of Human Genetics* 41: 566-83.
- Pedersen, F. M., O. Hamburg, K. Hess, and L. Ovesen. 1991. The effect of vitamin K on warfarin-induced anticoagulation. *Journal of Internal Medicine* 229: 517-20.
- Pettifor, J. M., and R. Benson. 1975. Congenital malformations associated with the administration of oral anticoagulants during pregnancy. *Journal of Pediatrics* 86: 459-62.
- Pietersma-de Bruyn, A. L., P. M. van Haard, M. H. Beunis, et al. 1990. Vitamin K₁ levels and coagulation factors in healthy term newborns till 4 weeks after birth. *Haemostasis* 20: 8-14.
- Poser, J. W., and P. A. Price. 1979. A method for decarboxylation of γ -carboxyglutamic acid in proteins. Properties of the decarboxylated γ -carboxyglutamic acid from calf bone. *Journal of Biological Chemistry* 254: 431-6.
- Preissner, K. T. 1990. Biological relevance of the protein C system and laboratory diagnosis of protein C and protein S deficiencies. *Clinical Sciences* 78: 351-64.
- Price, P. A. 1988. Role of vitamin K-dependent proteins in bone metabolism. *Annual Review of Nutrition* 8: 565-83.
- Price, P. A., and S. A. Baukol. 1980. 1,25-Dihydroxyvitamin D₃ increases synthesis of the vitamin K-dependent bone protein by osteosarcoma cells. *Journal of Biological Chemistry* 255: 11660-3.
- Price, P. A., M. R. Urist, and Y. Otawara. 1983. Matrix Gla protein, a new Γ -carboxyglutamic acid-containing protein which is associated with the organic matrix of bone. *Biochemical and Biophysical Research Communications* 117: 765-71.
- Quick, A. J. 1937. Coagulation defect in sweet clover disease and in hemorrhagic chick disease of dietary origin; consideration of the source of prothrombin. *American Journal of Physiology* 118: 260-71.
- Report of the Committee on Nutrition. 1961. Vitamin K compounds and the water-soluble analogues. *Pediatrics* 28: 501-7.
- Rick, M. E. 1990. Protein C and protein S. Vitamin K-dependent inhibitors of blood coagulation. *Journal of the American Medical Association* 263: 701-3.
- Second Report of the Sixty Plus Reinfarction Study Research Group. 1982. Risks of long-term oral anticoagulant therapy in elderly patients after myocardial infarction. *Lancet* 1: 64-8.
- Shah, D. V., and J. W. Suttie. 1974. The vitamin K-dependent in vitro production of prothrombin. *Biochemical and Biophysical Research Communications* 60: 1397-1402.
- Shaul, W. L., H. Emery, and J. G. Hall. 1975. Chondrodysplasia punctata and maternal warfarin use during pregnancy. *American Journal of Diseases of Children* 129: 360-2.
- Shearer, M. J., V. Allan, Y. Haroon, and P. Barkhan. 1980. Nutritional aspects of vitamin K in the human. In *Vitamin K metabolism and vitamin K-dependent proteins*, ed. J. W. Suttie, 317-27. Baltimore, Md.
- Shearer, M. J., A. Bach, and M. Kohlmeier. 1996. Chemistry, nutritional sources, tissue distribution and metabolism of vitamin K with special reference to bone health. *Journal of Nutrition* 126 (Supplement 4S): 1181S-6S.
- The Sixty Plus Reinfarction Study Research Group. 1980. A double-blind trial to assess long-term oral anticoagulant therapy in elderly patients after myocardial infarction. *Lancet* 2: 989-93.
- Smith, P., H. Arnesen, and I. Holme. 1990. The effect of warfarin on mortality and reinfarction after myocardial infarction. *New England Journal of Medicine* 323: 147-52.
- Soute, B. A. M., W. Muller-Ester, M. A. G. de Boer-van den Berg, et al. 1985. Discovery of a gamma-carboxyglutamic acid-containing protein in human spermatozoa. *FEBS Letters* 190: 137-41.
- Stenflo, J. 1976. A new vitamin K-dependent protein. Purification from bovine plasma and preliminary characterization. *Journal of Biological Chemistry* 251: 355-63.
1984. Structure and function of protein C. *Seminars in Thrombosis and Hemostasis* 10: 109-21.
- Stenflo, J., P. Fernlund, W. Egan, and P. Roepstorff. 1974. Vitamin K-dependent modifications of glutamic acid residues in prothrombin. *Proceedings of the National Academy of Sciences USA* 71: 2730-3.
- Suttie, J. W. 1985. Vitamin K-dependent carboxylase. *Annual Review of Biochemistry* 54: 459-77.
1990. Warfarin and vitamin K. *Clinical Cardiology* 13: VI-16-8.

1991. Vitamin K. In *Handbook of vitamins*, ed. L. J. Machlin, 145-94. New York.
1995. The importance of menaquinones in human nutrition. *Annual Review of Nutrition* 15: 399-417.
- Suttie, J. W., L. L. Mumma-Schendel, D. V. Shah, et al. 1988. Vitamin K deficiency from dietary vitamin K restriction in humans. *American Journal of Clinical Nutrition* 47: 475-80.
- Thorp, J. A., L. Gaston, D. R. Caspers, and M. L. Pal. 1995. Current concepts and controversies in the use of vitamin K. *Drugs* 49: 376-87.
- Townsend, C. W. 1894. The haemorrhagic disease of the newborn. *Archives of Pediatrics* 11: 559-65.
- Tulchinsky, T. H., M. M. Patton, L. A. Randolph, et al. 1993. Mandating vitamin K prophylaxis for newborns in New York State. *American Journal of Public Health* 83: 1166-8.
- Uotila, L. 1990. The metabolic functions and mechanism of action of vitamin K. *Scandinavian Journal of Clinical Laboratory Investigations* 50 (Supplement 201): 109-17.
- Usui, Y., N. Nishimura, N. Kobayashi, et al. 1989. Measurement of vitamin K in human liver by gradient elution high-performance liquid chromatography using platinum-black catalyst reduction and fluorimetric detection. *Journal of Chromatography* 489: 291-301.
- Vermeer, C., B. L. M. G. Gijsberg, A. M. Craciun, et al. 1996. Effects of vitamin K on bone mass and bone metabolism. *Journal of Nutrition* 126 (Supplement 4S): 1187S-91S.
- Vermeer, C., K.-S. G. Jie, and M. H. Knapen. 1995. Role of vitamin K in bone. *Annual Review of Nutrition* 15: 1-22.
- Vitamin K Ad Hoc Task Force, American Academy of Pediatrics. 1993. Controversies concerning vitamin K and the newborn. *Pediatrics* 91: 1001-3.
- Walker, F. J. 1980. Regulation of activated protein C by a new protein. A possible function for bovine protein S. *Journal of Biological Chemistry* 255: 5521-4.
- Tsai, K. S., H. Heath, R. Kumar, et al. 1984. Impaired vitamin D metabolism with aging in women. Possible role in the pathogenesis of senile osteoporosis. *Journal of Clinical Investigation* 73: 1668-72.
- Wachman, A., and D. S. Bernstein. 1968. Diet and osteoporosis. *Lancet* 1: 958-9.
- Walker, P. M., and H. M. Linkswiler. 1972. Calcium retention in the adult human male as affected by protein intake. *Journal of Nutrition* 102: 1297-1302.
- Warren, J. M., and H. Spencer. 1972. Analysis of stable strontium in biological materials by atomic absorption spectrophotometry. *Clinical Chimica Acta* 38: 435-9.
1976. Metabolic balances of strontium in man. *Clinical Orthopedics and Related Research* 117: 307-20.
- Wasserman, R. H., R. A. Corradina, and A. N. Taylor. 1968. Vitamin D-dependent calcium-binding protein. Purification and some properties. *Journal of Biological Chemistry* 243: 3978-86.
- Wasserman, R. H., and A. N. Taylor. 1968. Vitamin D-dependent calcium-binding protein. Response to some physiological and nutritional variables. *Journal of Biological Chemistry* 243: 3987-93.
- Watts, N. B., S. T. Harris, H. K. Genant, et al. 1990. Intermittent cyclical etidronate treatment of postmenopausal osteoporosis. *New England Journal of Medicine* 323: 73-9.
- Whedon, G. D. 1959. Effects of high calcium intake on bone, blood and soft tissue: Relationship of calcium intake to balance in osteoporosis. *Federation Proceedings* 18: 1112-18.
- WHO (World Health Organization). 1962. *Calcium requirements*. Technical Report Series, No. 230. Geneva.
- Woodhouse, N. J. Y., P. Bordier, M. Fisher, et al. 1971. Human calcitonin in the treatment of Paget's bone disease. *Lancet* 1: 1139-43.
- Woodhouse, N. J. Y., F. H. Doyle, and G. F. Joplin. 1971. Vitamin D deficiency and primary hyperparathyroidism. *Lancet* 2: 283-7.
- Woodhouse, N. J. Y., S. Tun Chot, P. Bordier, et al. 1973. Vitamin D administration in primary hyperparathyroidism. *Clinical Science* 44: 3P.
- Yates, A. J., R. O. C. Oreffo, K. Mayor, et al. 1991. Inhibition of bone resorption by inorganic phosphate is mediated by both reduced osteoclast formation and decreased activity of mature osteoclasts. *Journal of Bone and Mineral Research* 6: 473-8.
- Zipkin, I., A. S. Posner, and E. D. Eanes. 1962. The effect of fluoride on the X-ray-diffraction pattern of the apatite of human bone. *Biochimica Biophysica Acta* 59: 255-8.



IV.B

Minerals

IV.B.1 Calcium

This chapter deals with the history of calcium and its metabolism in adult humans. It should be read in conjunction with Chapter IV.D.4 on osteoporosis, which contains a further discussion of calcium requirements and the effects of a deficiency in adults, and with Chapter IV.A.4 on vitamin D, which deals with the history of rickets, a disease caused in part by a deficiency of calcium resulting from reduced intake or poor absorption.

Historical Aspects

Bone is the main depository of calcium. Fully 99 percent of the body's calcium is in the skeleton, with the rest in extracellular fluid and soft tissues. As early as the sixteenth century, it was recognized by a Dutch physician that the skeleton is not an inactive but a dynamic tissue under hormonal influence and capable of remodeling throughout life (Lutwak, Singer, and Urist 1974). Two specific types of bone cells acted upon in these processes are the osteoblasts (involved in bone formation) and the osteoclasts (involved in bone resorption).

Another important discovery in the history of calcium was made by Sidney Ringer more than 100 years ago. He demonstrated that the contractility of cardiac muscle was stimulated and maintained by the addition of calcium to the perfusion fluid (Ringer 1883). It has also been shown that this important effect of calcium is not limited to cardiac muscle but has a generalized, activating effect in practically all differentiated cells (Opie 1980; Rubin 1982; Campbell 1986). However, in addition to calcium, the presence of specific concentrations of sodium and potassium are needed to achieve this effect (Mines 1911; Lowenstein and Rose 1978).

The state of calcium in the body investigated many decades ago has been subsequently summarized ("The Classic" 1970). The amount of calcium in the body is greater than that of any other positively

charged mineral (1,160 grams [g]). Its place of storage is the skeleton, although the small amounts present in the compartments of the extracellular fluid and those of the soft tissues are of great physiological importance because the calcium stored in bone maintains equilibrium with other calcium pools in the body (Rubin 1982).

The interaction and balance of vitamin D, calcitonin, and parathyroid hormone (PTH) are the pillars of normal calcium metabolism, and they play an important role in maintaining a highly controlled calcium homeostasis and normal serum calcium level. This process is based on the direct or indirect action of these substances on the skeleton. Vitamin D is produced by ultraviolet radiation of sunlight on the skin, which yields the vitamin D precursor ergosterol (Hess and Weinstock 1924). The hormones calcitonin and PTH are created, respectively, by the C cells of the thyroid gland and by the parathyroid glands. A deficiency as well as an excessive production and secretion of these three substances can lead to the development of specific disease states.

Vitamin D

Vitamin D, discovered at the beginning of the third decade of the twentieth century, is the substance that acts to prevent rickets. Many of the original reports of the discovery of vitamin D have been republished (Robertson 1969; Pramanik, Gupta, and Agarwal 1971) and are cited here for those who wish to consult them. In addition, the early discovery of the physiological importance of calcium (Ringer 1883) has been redescribed (Fye 1984; Ebashi 1987). The chemical structure of vitamin D was identified by a German chemist, Adolf Windaus, in 1931, and the subsequent commercial production of vitamin D practically eliminated the occurrence of rickets – the vitamin D deficiency disease of children – although treatment with cod-liver oil (rich in vitamin D) also played an important role in the eradication of this disease.

Even though our basic knowledge of vitamin D was obtained between 1919 and 1924, it was not until the 1960s that the vitamin D metabolites were

discovered (Lund and DeLuca 1966; Ponchon and DeLuca 1969; Olson, Jr., and DeLuca 1973; DeLuca 1979, 1980, 1988). H. F. DeLuca postulated that vitamin D must be hydroxylated first in the liver and subsequently by 1-alpha-hydroxylation in the kidney to produce the vitamin D hormone 1-alpha-25-dihydroxyvitamin D₃, 1,25(OH)₂D₃. This is the important vitamin D metabolite that actively affects the absorption of calcium from the intestine. Further studies regarding the mechanism of calcium absorption from the intestine are credited to R. H. Wasserman, who discovered the vitamin D-dependent “calcium binding protein” (CaBP) in the duodenum (Kallfelz, Taylor, and Wasserman 1967; Wasserman, Corradina, and Taylor 1968; Wasserman and Taylor 1968).

Rickets, caused by vitamin D deficiency, was produced experimentally in the 1920s, and much was learned about this illness, which occurs in infancy and childhood at ages of rapid growth (Sherman and Pappenheimer 1921; Steenbock and Black 1924; Pettifor et al. 1978; Pettifor and Ross 1983). Rickets was common in the early part of the twentieth century, and more than 100 years ago it was recognized that a lack of sunlight was responsible, particularly in northern latitudes, where deprivation is most likely to occur. It was also noted that sunlight is curative (Palm 1890; Huldshinsky 1919; Mellanby 1921).

Symptoms of rickets include impaired calcification and excess formation of cartilage in areas of bone growth. The adult form of the disease is called osteomalacia, in which newly formed osteoid does not calcify. This occurs in persons who voluntarily or for other reasons are homebound and therefore not exposed to sunlight.

Several early investigators stated that both rickets and osteomalacia can also be caused by a nutritional deficiency of calcium (McCollum et al. 1921; Theiler 1976; McCollum et al. 1995). Although such a situation would be uncommon, it is possible that the theory was proposed because healing of rickets occurred when a low, insufficient calcium intake of barely more than 100 milligrams (mg)/day was raised to 1,200 mg/day with the simultaneous use of vitamin D. More recent investigators believed it unlikely that vitamin D deficiency alone could produce rickets or osteomalacia but that these diseases occur when there is coexistent calcium deficiency (Pettifor et al. 1981).

Calcitonin

The hormone calcitonin was discovered in 1962 by Harold Copp, who reported that it originates in the parathyroid glands (Copp et al. 1962; Copp 1967, 1969). Calcitonin decreases the calcium level in blood, and this effect was ascribed to decreased bone resorption. The first report of the use of calcitonin in humans came shortly after its discovery (Foster et al. 1966), and further studies revealed that this hormone actually originates in the C cells of the thyroid gland – which, for a while, gave rise to the use of the name

“thyrocalcitonin” (Hirsch and Munson 1969). It has been shown experimentally that calcitonin affects not only bone resorption but also bone formation (Baylink, Morey, and Rich 1969).

Because of the ability of calcitonin to cause a decrease in bone resorption, it has been utilized in the treatment of patients with Paget’s disease – a deforming and frequently disabling bone disease diagnosed in England more than a century ago (Paget 1877; Bijvoet, Van der Sluys Veer, and Jansen 1968; Shai, Baker, and Wallach 1971; Woodhouse Bordier, et al. 1971). Two available types of calcitonin, primarily salmon calcitonin but also porcine calcitonin, have been joined by human calcitonin. All three types have been (and still are) used in investigative studies of their comparative effectiveness in treating Paget’s disease.

Parathyroid Hormone

Parathyroid hormone (PTH), discovered in the early 1920s (Collip 1925), consists of four small pea-sized structures, two of which are located on the upper pole and two at the lower pole of each thyroid lobe. When parathyroid glands, which produce PTH, are functioning normally, the level of secretion depends on the serum calcium level, which in turn depends, in part, on the fraction of the dietary calcium that is absorbed from the intestine and the amount of calcium released from the skeleton by bone resorption. The dietary contribution of calcium to the serum calcium level influences the extent of bone resorption that is induced by PTH to maintain the well-controlled homeostasis of the normal serum calcium level.

The most common clinical aberration of calcium metabolism in hyperparathyroidism is a high level of serum calcium, a low level of serum phosphorus, and frequently, but not invariably, elevated levels of the serum enzyme alkaline phosphatase. There may also be evidence of bone loss on roentgenograms (which can mimic osteopenia) and, more rarely, cystic bone lesions. Kidney stone formation and peptic ulcer of the stomach may also result.

When there is hyperfunction of any of the four parathyroid glands, and excess PTH is secreted, a pathological condition develops that is called primary hyperparathyroidism. This endocrine disorder that affects bone metabolism (primarily the metabolism of calcium and phosphorus) was discovered in the early 1920s in Vienna and discussed in print two years later (Mandl 1926). In 1925, the condition was observed in the United States (Hannon et al. 1930), and shortly after, the first series of patients with hyperparathyroidism in the United States was described (Albright, Aub, and Bauer 1934). A classic study of the parathyroid glands followed in the late 1940s (Albright and Reifenstein 1948).

The symptoms and the treatment of hyperparathyroidism have been extensively discussed in older medical textbooks, which state, as did the classic arti-

cle by F. Albright, J. C. Aub, and W. Bauer in 1934, that the cause of primary hyperparathyroidism is the enlargement of one or more of the parathyroid glands to form a benign tumor (an adenoma) or a multiple adenomata. However, we now know that the adenoma of a single parathyroid gland or multiple adenomata of the parathyroid glands are the result of hyperfunction of the parathyroid glands - the increased secretion of PTH. The conventional treatment for primary hyperparathyroidism is the surgical removal of the enlarged parathyroid gland or glands.

Reports in the literature also attest to the importance of the role of vitamin D in normal parathyroid function (Mawer et al. 1975). There are also reports of vitamin D deficiency in hyperparathyroidism (Woodhouse, Doyle, and Joplin 1971; Lumb and Stanbury 1974; Mawer et al. 1975; Stanbury 1981), and the use of vitamin D has been recommended for the medical treatment of patients with primary hyperparathyroidism (Woodhouse et al. 1973).

Nutritional Aspects

The nutritional importance of calcium for normal growth and maintenance has been recognized since studies were carried out with experimental animals several decades ago. Those studies that focused on the human calcium requirement, for example, were conducted as early as the 1920s (Sherman 1920), and the importance of calcium in the human metabolism (and its use in therapy) was recognized in the early 1930s (Cantarow 1933). Subsequently, it was shown that a nutritionally high calcium intake resulted in an increased growth rate of children (Stearns, Jeans, and Vandecar 1936; Jeans and Stearns 1938). Textbooks on the importance of calcium began to appear by the late 1930s (Shohl 1939), and the effect of dietary factors on the intestinal absorption of calcium was examined in the early 1940s in a study demonstrating that the presence of dietary protein is necessary for the intestinal absorption of calcium (McCance, Widdowson, and Lehmann 1942).

Early in the 1950s, the interaction in humans of dietary calcium with phosphorus was investigated (Leichsenring et al. 1951), and at about the same time, two studies of the human calcium requirement were carried out in prisons, one in Peru (Hegsted, Moscoso, and Collazos 1952) and one in Scandinavia (Malm 1958). Both of these investigations reported a very low calcium requirement of about 250 mg/day. However, shortly thereafter, it was made clear that a calcium intake of 1,000 mg is necessary to achieve calcium balance (Whedon 1959).

The calcium requirement for skeletal maintenance was investigated at the beginning of the 1970s (Garn 1970), and this study was followed by a determination of the calcium requirement both for middle-aged women (Heaney, Recker, and Saville 1977) and for elderly people (Heaney et al. 1982). In 1984, a strictly

controlled metabolic study revealed that 800 mg/day of calcium is insufficient for middle-aged men and that an intake of 1,200 mg is desirable (Spencer, Kramer, Lesniak, et al. 1984). This recommendation was subsequently adopted at a National Institutes of Health Consensus Conference (1995).

As hinted at by the prison studies already mentioned, some investigators believed that humans can adapt to a long-term low calcium intake. Experimentally, however, this was not found to be the case at a calcium intake of 200 mg/day in adults (Spencer and Kramer 1985). It is important to note that bioavailability of calcium for absorption is greatly reduced by substances such as fiber (in large amounts) and phytate in food (Harrison and Mellanby 1934; Mellanby 1949; Ismail-Beigi et al. 1977; Cummings 1978). The negative effect of phytate on calcium has been extensively investigated (Reinhold et al. 1973), but it has also been shown that calcium bioavailability depends on the availability of trace minerals such as copper, iron, and manganese (Strause et al. 1994), and that deficiencies of these trace elements decrease the calcium concentration in bone. The trace element zinc, for example, affects intestinal absorption of calcium in humans (Spencer et al. 1987; Spencer, Norris, and Osis 1992).

The impact of other substances - such as lactose - on the absorption of calcium was reported as early as 1940 (Mills et al. 1940), and 20 years later, a follow-up report was published (Greenwald, Samachson, and Spencer 1963). An extensive review of human calcium requirements was published in the early 1970s (Irwin and Kienholz 1973), and other such surveys have subsequently appeared (Bronner and Peterlik 1995; Bronner and Stein 1995). The suggested recommended dietary allowance (RDA) of calcium in the United States is determined by the Food and Nutrition Board of the National Academy of Science.

Osteoporosis

Osteoporosis is the most common systemic bone disorder in the United States, affecting an estimated 20 million women and perhaps 10 million men. The incidence of osteoporosis is highest for women in the decade after menopause. It occurs much later in men - between the seventh and eighth decades of life. The high incidence of osteoporosis is a major public health problem because of serious complications that may require extensive and costly medical care. The bone loss in women at middle age and older is the result of hormonal deficiency (the loss of estrogen), aging, and other factors that appear to play an important role in causing - or intensifying - the loss of calcium from the skeleton. It has become more and more evident that sufficient calcium intake over the years can play a major role in preventing osteoporosis in later life (Matkovic et al. 1979; Sandler et al. 1985).

The disease develops insidiously over many years, is often asymptomatic, and is frequently discovered

accidentally during radiological examination for unrelated medical conditions (Spencer and Kramer 1987). Unfortunately, this means that treatment of osteoporosis is frequently delayed until it is quite advanced, and it may not be detected until complications such as skeletal fractures arise. In the past, physicians assumed that there was no effective treatment available for this disorder. However, in recent years, effective therapeutic modalities have become available for the treatment and even for the prevention of the disease. Among the newer treatment modalities are the biphosphonates (Fogelman et al. 1986; Francis 1995).

When the diagnosis of bone loss is established roentgenographically, osteoporosis is usually advanced, because 30 percent of bone mineral must be lost before the loss will show up on conventional skeletal X rays. The vertebral bodies show poor mineralization and become biconcave; adjacent vertebrae have a “fish-mouth” appearance; and there may be wedging of the involved vertebrae. These changes in bone structure are usually associated with loss of body height, kyphosis, deformity of the chest, and the presence of a “dowager’s hump.” Routine X rays of the skeleton are not reliable indicators of bone density because of differences in techniques and in subjective interpretations of X-ray findings.

When demineralization of bone is seen on an X ray, this finding does not differentiate between osteoporosis and other demineralizing bone diseases, such as multiple myeloma or hyperparathyroidism. Reliable methods – such as bone density measurements using single and double photon absorptiometry and determining the cortical index – are now available for diagnosing osteoporosis earlier and with more precision than the use of conventional radiographs offers. However, photon absorptiometry (Mazess et al. 1988) is not routinely done because the equipment is not available in all medical centers. Cortical index measurements can be determined by analyzing X rays of the hand and by relating the cortical thickness to the total width of the metacarpal bone.

Another method of determining bone density is radiographic absorptiometry, which utilizes radiographs of the hands and a computer scanner. Newer, sophisticated methods of analyzing bone mass are available in specialized centers. In general, X rays of the thoracic and lumbar spine can indicate the potential existence of postmenopausal osteoporosis. Bone biopsies (needle biopsies) can establish the differential diagnosis of osteoporosis, but this technique is invasive and not acceptable to many patients.

As already noted, a low calcium intake, over prolonged periods of time, results in calcium loss from the skeleton and in a negative calcium balance (Table IV.B.1.1). This condition has an adverse effect on the maintenance of the normal bone structure, and a continued loss of calcium from the skeleton results from bodily efforts to maintain a normal serum calcium level. Moreover, a low calcium intake stimulates the

Table IV.B.1.1. *Calcium balances of males and females during a low calcium intake*

Number of patients	Type of patient	Study days	Calcium, mg/day			
			Intake	Urine	Stool	Balance
10	Male	36	214	104	196	-86
10	Female	30	192	92	174	-74

parathyroid glands to excess secretion of PTH, which in turn leads to increased bone resorption in order to maintain the homeostasis of the normal serum calcium level.

The retention of calcium from a high calcium intake, however, is frequently lower for those with osteoporosis than it is for nonosteoporotic subjects or for patients who suffer from other conditions of bone loss, such as hyperparathyroidism or hyperthyroidism. The absorption of calcium from the intestine is decreased in elderly persons, including osteoporotic females. This appears to be caused in part by a relative vitamin D deficiency that occurs with aging (Tsai et al. 1984) but may also result from generalized functional changes of the intestinal mucosa with age, relating to the absorbability of nutrients in general.

Yet even when calcium is administered by the intravenous route (bypassing the intestine), the retention of calcium is low in osteoporotic patients, indicating the inability of the skeleton to accept the added calcium that has entered the circulation directly (Spencer, Hausinger, and Laszlo 1954). Blood levels of the active vitamin D metabolite $1,25(\text{OH})_2\text{D}_3$ have been shown to be low in these patients (Tsai et al. 1984). In the differentiation of calcium loss in conditions other than osteoporosis, the determination of urinary hydroxyproline and of free cortisol levels may be helpful.

Although calcium is the major mineral of the bone structure, other minerals – such as phosphorus, magnesium, zinc, and fluoride – are also important because of the interaction of calcium with these elements and their specific functions in the body. Before these interactions are considered, however, the calcium requirement – the amount of calcium needed to maintain a normal calcium status – requires some attention. The data employed in the following discussion were derived from studies carried out under strictly controlled dietary study conditions in the Metabolic Research Unit at the Veterans Administration (VA) Hospital at Hines, Illinois.

The Calcium Requirement

An adequate intake of calcium throughout life has been shown to play an important role in maintaining the normal bone structure and to contribute to the peak bone mass that is achieved at between 25 and 30 years of age (Heaney 1982; Heaney et al. 1982). The bone mass begins to decline after age 35, and this

decrease accelerates in females after the menopause. One can safely assume that the adverse and deleterious effects of aging on the bone mass and on the bone structure would be diminished in advancing age if the skeleton could be more robust at the time when the inevitable and accelerated bone loss begins.

In view of these considerations, the questions arise: What should be the daily calcium intake for the elderly? Is the RDA for calcium adequate for this age group (National Research Council 1989)? For calcium, the RDA for young persons up to the age of 24 years is 1,200 mg, but it is reduced to 800 mg/day for all age groups after age 25, including elderly women (National Research Council 1989). It therefore has seemed important to examine whether the 800 mg calcium intake is adequate to maintain a normal calcium balance at middle age (and to prevent excessive calcium and bone loss) and thereby maintain a normal calcium status and a normal skeletal structure in women with advancing age.

Our studies showed that the calcium balance of middle-aged males was only slightly positive, +30 mg/day with an intake of 800 mg calcium (Table IV.B.1.2) – without considering the dermal loss of calcium, which is usually not determined in metabolic balance studies. The latter is stressed because in calcium balance studies, only the urinary and fecal calcium excretions are determined, and the sum of these excretions is related to the calcium intake. A large percentage of these fully ambulatory middle-aged males (34 percent) were in negative calcium balance at the 800 mg/day intake level. Increasing the calcium intake from 800 to 1,200 mg/day resulted in a significant increase of the calcium balance (Spencer, Kramer, Lesniak, et al. 1984). Yet adding another 800 or even 1,100 mg calcium to the 1,200 mg intake did not significantly improve the calcium balance (Table IV.B.1.2). Such a plateau of the calcium balance at the 1,200 mg calcium intake indicates a threshold for the intestinal absorption of calcium at this level (Spencer and Kramer 1987).

Table IV.B.1.2 also shows that urinary calcium was the same whether the calcium intake was 800, 1,200, or 2,000 mg/day. Only when the calcium intake was increased further – to 2,300 mg/day – did urinary calcium increase. This point is emphasized because of the widespread – and unjustified – concern that increasing the calcium intake beyond the 800 mg level may result in kidney stone formation. Such a concern does become important, however, if there is a history of kidney stone formation or even a family history of renal stones, because a certain percentage of kidney stone formers are hyperabsorbers of calcium.

Although these findings came from studies carried out in an all-male VA hospital, there is no reason to assume that they would not also apply to females of similar age (average 54 years) or older. Indeed, in view of the large percentage of the male subjects with negative calcium balances at the 800 mg calcium intake

Table IV.B.1.2. *Studies of the calcium requirement*

Calcium intake, mg/day	Calcium, mg/day			
	Intake	Urine	Stool	Balance
200	250	85	265	-100
800	820	180	610	+30
1,200	1,250	170	980	+100
2,000	2,060	175	1,750	+135
2,300	2,350	245	1,945	+160

Note: Calcium intake greater than 250 mg was due to the addition of calcium gluconate tablets to the constant low-calcium diet.

level, and in view of the known calcium loss suffered by middle-aged and older women, it seems clear that the calcium intake of older women should be at least as great as that of considerably younger persons – that is, 1,200 mg/day (National Research Council 1989). Thus, some have recommended that the calcium intake of postmenopausal women should be increased to 1,500 mg/day (Heaney 1982).

It is also important to consider that although calcium intake may be adequate, there are certain factors that influence the utilization of calcium. The following subsections describe the effects of various minerals, nutrients, and drugs – phosphorus, magnesium, fluoride, zinc, strontium, protein, alcohol, and medications – on the metabolism of calcium.

Phosphorus

The mineral phosphorus is present in practically all cells of the body and is closely linked to the metabolism of both calcium and protein. The main storehouse of phosphorus is the skeleton, and the bone crystal hydroxyapatite consists of calcium phosphate. But significant amounts of phosphorus are also contained in the soft tissues. Calcium cannot be retained by itself in bone but is retained together with phosphate. In fact, in calcium balance studies, the retention of calcium improved during high phosphorus intakes (Spencer et al. 1965; Spencer, Kramer, Osis, et al. 1978a).

These observations were made during a high phosphorus intake of 2,000 mg/day, compared with a phosphorus intake of 800 mg/day in the control study. But although the phosphorus intake was increased by a factor of approximately 2.5 (from 800 to 2,000 mg/day) in these investigations, no adverse effect of phosphorus on calcium absorption or calcium balance was observed, regardless of whether the high phosphorus intake was given during a very low calcium intake of 200 mg/day or during a high calcium intake of 2,000 mg/day (Spencer et al. 1965; Spencer, Kramer, Osis, et al. 1978a).

With a high phosphorus intake, fecal calcium was slightly but not significantly increased. Therefore, the intestinal absorption of calcium, determined with tracer doses of ^{47}Ca , was also not decreased (Spencer

Kramer, Osis, et al. 1978a). This result is in contrast to the general unjustified belief that a high phosphorus intake decreases humans' intestinal absorption of calcium. Such an assumption is based primarily on animal studies (Draper, Sie, and Bergan 1972; LaFlamme and Jowsey 1972).

The difference in effect, however, points out the difficulty and unreliability of extrapolating animal data to humans. Very few, if any, strictly controlled dietary studies of the effect of phosphorus on calcium absorption have been carried out in humans, and it is possible that amounts of phosphorus greater than 2,000 mg/day may have adverse effects on calcium absorption. The recently reported beneficial effect of biphosphonates in the treatment of osteoporosis (Watts et al. 1990) may result from the action of phosphate in inhibiting bone resorption by decreasing the activity of the osteoclasts (the bone-reabsorbing cells) (Hodsman 1989). Inorganic phosphate has been reported to have this effect (Yates et al. 1991). The usual dietary phosphorus intake is approximately 1,200 mg/day, most of it from proteins in dairy products and meat.

The main impact of added phosphorus on calcium metabolism is a consistent decrease in urinary calcium, regardless of whether the high phosphorus intake occurs during a low or high calcium intake (Goldsmith et al. 1967; Goldsmith et al. 1969; Spencer, Kramer, Osis, et al. 1978a). The decrease in urinary calcium during a high phosphorus intake is usually believed to reflect decreased calcium absorption. However, as already noted, the intestinal absorption of calcium did not decrease during a high phosphorus intake (Spencer, Kramer, Osis, et al. 1978a). Although there was a slight increase in fecal calcium during the high phosphorus intake in our studies (Spencer et al. 1965; Spencer, Kramer, Osis, et al. 1978a), this increase was not significant.

Several factors other than a decrease in calcium absorption can contribute to the decrease in urinary calcium during a high phosphorus intake; these include decreased bone resorption, increased mineralization, and increased bone formation (Pechet et al. 1967; Flanagan and Nichols 1969). As the intestinal absorption of calcium did not decrease during differ-

ent high phosphorus intakes up to 2,000 mg/day, the dietary Ca:P ratio does not appear to play an important role in the intestinal absorption of calcium. This viewpoint was also expressed in 1962 by the World Health Organization.

The importance of phosphorus in human health and in calcium metabolism is indicated by the deleterious effects of phosphorus depletion, which is most commonly induced by the use of medications such as aluminum-containing antacids. Even relatively small doses of such antacids induce a considerable loss of calcium (Spencer, Kramer, Norris, et al. 1982). In addition, small doses – as well as larger therapeutic doses (Lotz, Zisman, and Bartter 1968) – induce phosphorus depletion by the complexation of phosphate through aluminum in the intestine. This is evidenced by a very significant increase in fecal phosphorus, which may be as great as the entire dietary phosphorus intake (Table IV.B.1.3).

The loss of phosphorus via the intestine may lead to its removal from bone in order to maintain the phosphorus level in tissues, enzymes, and plasma. The removal of phosphorus from bone appears to be associated with simultaneous removal of calcium from the skeleton, resulting in an increase in urinary calcium and a negative calcium balance. Thus, significant bone loss has been observed in patients who have taken commonly used aluminum-containing antacids for prolonged periods of time.

The prolonged use of intravenous fluids in the absence of food intake may also result in phosphorus depletion, which is associated with clinical symptoms of weakness and fatigue. To our knowledge, no data are available on the relationship of this type of induced phosphorus depletion and calcium metabolism.

Magnesium

Magnesium is an essential nutrient of great importance in controlling normal cardiac rhythm and cardiovascular function. A low magnesium status has been associated with cardiac arrhythmia as well as with cerebrovascular spasm (Altura and Altura 1981). The RDA for magnesium is 300 mg for women and 400 mg for men (National Research Council 1989). Bone is an important repository for magnesium.

Table IV.B.1.3. *Effect of aluminum-containing antacids on the calcium and phosphorus balance*

Patient	Study	Calcium, mg/day				Phosphorus, mg/day			
		Intake	Urine	Stool	Balance	Intake	Urine	Stool	Balance
1	Control	231	89	229	-87	845	533	323	-11
	Maalox ^a	225	123	264	-132	862	138	673	+51
2	Control	254	86	240	-72	933	482	299	+152
	Maalox ^b	279	421	380	-522	873	18	890	-35

^aAluminum-magnesium hydroxide, 30 ml three times daily.

^bEvery hour while awake (15 doses = 450 ml per day).

Magnesium balance studies have shown great variability of the magnesium balance in different individuals, regardless of magnesium intake. During a relatively low magnesium intake of about 220 mg/day, the magnesium balance is usually negative. However, equilibrium or even positive magnesium balances have been observed during a low magnesium intake in our studies on male subjects (Spencer, Lesniak, Kramer, et al. 1980).

The intestinal absorption of dietary magnesium is approximately 50 percent of its intake in persons with normal renal function, regardless of calcium intake (Spencer, Schwartz, and Osis 1988). In contrast, patients with chronic renal failure absorb considerably less magnesium, about one-third that of those with normal renal function (Spencer, Lesniak, Gatz, et al. 1980). It appears that magnesium balance depends on both the present and the past magnesium status of the individual. This was demonstrated by increasing the magnesium intake in adequately nourished subjects from a low of 220 to a relatively high 800 mg/day. Such an increase, however, did not improve magnesium balance when compared with intakes of 220 to 300 mg/day. It would appear, therefore, that the diet of these patients prior to the magnesium absorption studies contained an adequate amount of magnesium.

The interaction of magnesium with calcium warrants discussion because magnesium has been reported to have variable effects on the intestinal absorption of calcium. Reports of some animal studies have suggested that magnesium decreases the intestinal absorption of calcium (O'Dell 1960). But other studies have indicated no change in humans (Schwartz et al. 1973), and still others have reported an increase in calcium absorption. Our investigations under controlled dietary conditions have shown that increasing the magnesium intake more than threefold - from 220 to 860 mg/day - had no effect on the intestinal absorption of calcium (Spencer et al. 1994).

Changing the focus to the effect of calcium on magnesium metabolism, we find reports of the intestinal absorption of magnesium being impaired by calcium in animals (Morris and O'Dell 1961). By contrast, however, our studies on humans - employing both magnesium balance studies and tracer studies using radioactive magnesium (Mg^{28}) - have conclusively shown that both intestinal absorption of magnesium and the magnesium balance did not change when the calcium intake was increased from 200 to 800 and even to 2,000 mg/day (Schwartz et al. 1973).

As phosphorus and calcium are closely linked, the effect of phosphorus on the metabolism of magnesium might warrant brief mention. Whereas phosphorus, like calcium, decreased the absorption of magnesium in animals (O'Dell 1960), increased amounts of phosphorus - up to 2,000 mg/day - had little effect on the magnesium balance of humans (Spencer, Lesniak, Kramer, et al. 1980), regardless of whether these studies were carried out during a low or high calcium intake. How-

ever, a high magnesium intake led to increased fecal phosphorus excretion and to a less positive or even negative phosphorus balance, probably because of the formation of magnesium-phosphate complexes in the intestine. But despite this effect, the calcium balance was not affected (Spencer et al. 1994). Further studies of the effect of magnesium on phosphorus metabolism in humans are needed. Such studies should seek to determine at which level a high magnesium intake has adverse effects on phosphorus metabolism and, potentially therefore, on calcium metabolism.

Fluoride

Fluoride enters the human food chain because of the fluoride content of water and soil. The skeleton is its major repository in the body. Not only is fluoride beneficial in preventing dental caries in children, but there is evidence that it is also important in maintaining the normal bone structure (Zipkin, Posner, and Eanes 1962). Consequently, fluoride affects the metabolism of calcium. Surveys in the United States and in Finland have demonstrated that the prevalence of osteoporosis is lower in places where the water is naturally high in fluoride content than in areas where this condition does not obtain (Bernstein, Sadowsky, and Hegsted 1966).

Because the main storehouse of fluoride is the skeleton, where it is incorporated in the bone crystal hydroxyapatite (leading to its increased strength) (Zipkin et al. 1962), fluoride has been used for the treatment of osteoporosis since the early 1960s (Rich, Ensink, and Ivanovich 1964; Spencer et al. 1970). As fluoride may interact with various minerals, such as calcium, phosphorus, and magnesium, we investigated the human metabolism of fluoride during the intake of these inorganic elements. During a fluoride intake as high as 45 mg/day (taken as sodium fluoride), the single effect on calcium metabolism was a decrease in urinary calcium, whereas no change occurred in fecal calcium, the intestinal absorption of calcium, and endogenous fecal calcium (the amount of the absorbed calcium excreted into the intestine) (Spencer et al. 1970). There was also little change in the calcium balance, which depended on the decrease in urinary calcium during the high fluoride intake. Increasing phosphorus intake by a factor of 2.5, from 800 to 2,000 mg/day, had no effect on the fluoride balance, regardless of calcium intake (Spencer et al. 1975). Also, increasing magnesium intake approximately threefold during a high fluoride intake had no effect on the fluoride balance (Spencer, Kramer, Wiattowski, et al. 1978). Several biochemical and therapeutic aspects of the fluoride metabolism in humans have been summarized (Spencer, Osis, and Lender 1981).

Zinc

The importance of the trace element zinc in human health and nutrition has been emphasized in recent decades. The RDA of zinc is 15 mg/day for men and

12 mg/day for women (National Research Council 1989). Zinc and calcium appear to have the same binding sites in the intestine, and, therefore, the absorption and utilization of one may be inhibited by the other. Animal studies have shown that calcium inhibits the intestinal absorption of zinc (Hoekstra et al. 1956; Luecke et al. 1957), but our studies on humans demonstrated that calcium had no effect on the absorption of zinc (Spencer, Kramer, Norris, et al. 1984). Similarly, phosphorus – used in amounts of up to 2,000 mg/day, alone or combined with the same amount of calcium – did not affect the zinc balance nor the net absorption of zinc (Spencer, Kramer, Norris, et al. 1984). These findings contrasted with those obtained in animal studies, with the different results apparently arising from differences in body weights and differences in the amounts of calcium and phosphorus relative to body weight.

Another question was whether zinc affects the intestinal absorption of calcium. Our calcium absorption studies (using ^{47}Ca as the tracer) have conclusively shown that large doses of zinc (140 mg/day of zinc sulfate) significantly decreased the intestinal absorption of calcium when zinc supplements were given during a low calcium intake of 230 mg/day – but not during a calcium intake of 800 mg/day (Spencer et al. 1987). Further investigations have delineated the dose of zinc and the level of calcium intake at which the decrease of calcium absorption would not occur. Decreasing zinc intake from 140 mg to 100 mg during a low calcium intake of 230 mg/day, and decreasing calcium intake from 800 mg to 500 mg during the high zinc intake of 140 mg/day, in both cases had no adverse effect on the intestinal absorption of calcium (Spencer et al. 1992).

Zinc supplements are freely available “over the counter” and are used in unknown dosages by the public with equally unknown intakes of calcium. This practice is of concern for elderly women who may already have a low calcium status and yet may use large doses of zinc during a low calcium intake. Such a combination would decrease the intestinal absorption of calcium and thereby contribute to further deterioration of the state of calcium metabolism in these individuals.

Strontium

Little attention is being paid to strontium, which, like calcium, is primarily deposited in bone. Our studies have shown that the average dietary intake of strontium is low, at 1.5 to 2.5 mg/day (Warren and Spencer 1972, 1976). In 1950, it was suggested that strontium used in conjunction with calcium would be effective in therapy for osteoporosis and that the retention of both calcium and strontium would be additive (Shorr and Carter 1950). About 30 years later, other investigators again demonstrated the beneficial effect of strontium as a therapeutic agent for osteoporosis (Marie et al. 1985). Moreover, our investigations indicated that large amounts of elemental strontium, such as daily doses of

600 to 900 mg (taken as strontium lactate), were well retained and well tolerated. However, it was found that after discontinuation of strontium supplements, a very high percentage of the retained strontium was excreted in three to four weeks (Warren and Spencer 1976).

It may be speculated that this loss of strontium might not have occurred if calcium intake had been high. Some of our preliminary studies show that the intake of added strontium increases the intestinal absorption of calcium, which could be a significant finding. Further examinations of the effect of strontium on the absorption of calcium are needed because very few substances are known to increase the intestinal absorption of calcium.

Protein

Although proteins are not the same sort of nutrients as minerals, the relationship between proteins and the metabolism of calcium is important because proteins play a major role in the formation of the bone matrix and, conversely, have also been reported to bring about an increase in urinary calcium. However, such findings of calcium loss are primarily based on studies using purified proteins (Walker and Linkswiler 1972; Schuette, Zemel, and Linkswiler 1980). Calcium loss does not occur when complex proteins – which are part of the human diet – are used (Spencer, Kramer, Wiatrowski, et al. 1978; Spencer et al. 1983; Spencer, Kramer, and Osis 1988). The widespread belief that “proteins” generally are a cause of calcium loss frequently stems from reports and statements that do not specify or identify the type of protein used. But in considering the calciuric effect of protein in causing calcium loss, it is important to define the source and type of protein: Are the proteins isolated protein fractions, such as specific amino acids, or are they part of complexes of other nutrients, as is usually the case in the human diet?

Our studies, carried out under strictly controlled dietary conditions, have conclusively shown that a high protein intake (using red meat as its source, in large amounts of up to 550 g of meat per day) did not increase urinary calcium excretion (Spencer, Kramer, Osis, et al. 1978b; Spencer et al. 1983). Such a diet was given daily for as long as four months, and an example of the effect of this type of dietary protein on urinary calcium is shown in Table IV.B.1.4.

Moreover, during this high protein intake, there was no change in fecal calcium, nor in calcium balance, nor in the intestinal absorption of calcium, as determined in ^{47}Ca absorption studies. As already discussed, phosphorus decreases urinary calcium excretion. Dietary protein sources such as meat, milk, and cheese have a high phosphate content, which may explain why urinary calcium does not increase during the intake of these complex proteins. Therefore, the high phosphorus content of complex proteins may prevent and/or counteract any increase in urinary calcium – even with the consumption of red meat, an

Table IV.B.1.4. *Effect of a high-protein diet on calcium metabolism*

Patient	Study	Study days	Calcium, mg/day			
			Intake	Urine	Stool	Balance
1	Control	36	824	111	690	+23
	High protein	36	846	118	683	+45
2	Control	60	824	173	661	-10
	High protein	36	875	157	677	+42

Note: High protein given as red meat; high protein = 2 g/kg body weight compared with 1 gm/kg in the control study.

acid-ash food that would be expected to increase the urinary excretion of calcium (Wachman and Bernstein 1968).

Effects of Excessive Alcohol Consumption

Alcohol cannot be classified as a nutrient. However, chronic alcoholism can lead to bone loss (Feitelberg et al. 1987; Laitinen and Valimaki 1991) and to the development of osteoporosis. The etiology of osteoporosis in this case is multifactorial, but a major cause may well be the poor diet - especially low intakes of calcium, proteins, and vitamin D - associated with prolonged excessive alcohol consumption. Studies have suggested abnormalities in vitamin D metabolism (Gascon-Barre 1985) as well as in the adrenal function (Mendelson, Ogata, and Mello 1971), both of which affect the metabolism of calcium.

Changes in the pancreatic function in chronic alcoholism also lead to a loss of calcium - in this case because of the complexation of calcium with fat in the abdominal cavity - and there may be other factors not yet identified. Table IV.B.1.5 shows the prevalence of osteoporosis in our patients with chronic alcoholism (Spencer et al. 1986). Thirty-one percent were less than 45 years old, and 50 percent of these relatively young patients suffering from chronic alcoholism and osteoporosis were less than 40 years old (Spencer et al. 1986).

Table IV.B.1.5. *Patients with chronic alcoholism and osteoporosis*

Number of patients ^a	Age group, in years	Patients with osteoporosis	Percentage of patients with osteoporosis
5	20-30	-	-
33	31-45	14 ^b	31
58	46-62	31	69

^aAll male patients.

^bFifty percent of these patients were less than 40 years old.

Table IV.B.1.6. *Effect of corticosteroids on the calcium balance*

Patient	Study	Calcium, mg/day			
		Intake	Urine	Stool	Balance
1	Control	240	185	187	-132
	Corticoids ^a	234	287	177	-230
2	Control	1,955	90	1,733	+132
	Corticoids ^b	1,950	209	1,877	-136

^aAristocort, 20 mg/day for 24 days.

^bAristocort, 40 mg/day for 60 days.

Effect of Medications

Several medications affect the metabolism of calcium, primarily by increasing urinary calcium and thereby causing calcium loss. The effect of glucocorticoids in causing bone loss and osteoporosis, regardless of gender and age, is well known (Lukert and Adams 1976). Table IV.B.1.6 shows two examples of negative calcium balances during treatment with corticosteroids. These medications increase not only urinary but also fecal calcium, resulting in markedly negative calcium balances. This result occurs during both a low calcium intake of approximately 240 mg/day and an approximate tenfold increase of this amount. The data suggested that the loss of calcium induced by glucocorticoids was dose-dependent.

Herta Spencer

Bibliography

- Albright, F., J. C. Aub, and W. Bauer. 1934. Hyperparathyroidism. *Journal of the American Medical Association* 102: 1276-87.
- Albright, F., and E. C. Reifenstein. 1948. *The parathyroid glands and metabolic bone disease*. Baltimore, Md.
- Altura, B. M., and B. T. Altura. 1981. Magnesium ions and contraction of vascular smooth muscles: Relationship to some vascular diseases. *Federation Proceedings* 40: 2672-9.
- Avioli, L. V., J. E. McDonald, and S. W. Lee. 1965. The influence of age on the intestinal absorption of ⁴⁷Ca in women and its relation to ⁴⁷Ca absorption in postmenopausal osteoporosis. *Journal of Clinical Investigation* 44: 1960-7.
- Baylink, D., E. Morey, and C. Rich. 1969. Effect of calcitonin on the rates of bone formation and resorption in the rat. *Endocrinology* 84: 261-9.
- Bernstein, D., H. Sadowsky, and D. M. Hegsted. 1966. Prevalence of osteoporosis in high and low-fluoride areas in North Dakota. *Journal of the American Medical Association* 193: 499-504.
- Bijvoet, O. L. M., J. Van der Sluys Veer, and A. P. Jansen. 1968. Effects of calcitonin on patients with Paget's disease, thyrotoxicosis, or hypercalcaemia. *Lancet* 1: 876-81.
- Bronner, F., and M. Peterlik. 1995. International conference on progress in bone and mineral research. *Journal of Nutrition* 125: 1963S-2038S.

- Bronner, F., and W. D. Stein. 1995. Calcium homeostasis – an old problem revisited. *Journal of Nutrition* 125: 1987S-95S.
- Campbell, A. K. 1986. Lewis Victor Heilbrunn. Pioneer of calcium as an intracellular regulator. *Cell Calcium* 7: 287-96.
- Cantarow, A. 1933. *Calcium metabolism and calcium therapy*. Second edition. Philadelphia, Pa.
- The classic: The state of calcium in the fluids of the body. I. The conditions affecting the ionization of calcium. 1970. *Clinical Orthopaedics and Related Research* 69: 4-27.
- Collip, J. B. 1925. The extraction of a parathyroid hormone which will prevent or control parathyroid tetany and which regulates the level of blood calcium. *Journal of Biological Chemistry* 63: 395-438.
- Copp, D. H. 1967. Hormonal control of hypercalcemia. Historic development of the calcitonin concept. *American Journal of Medicine* 43: 648-55.
1969. Calcitonin and parathyroid hormone. *Annual Review of Pharmacology* 9: 327-44.
- Copp, D. H., E. C. Cameron, B. A. Cheney, et al. 1962. Evidence for calcitonin – a new hormone from the parathyroid that lowers blood calcium. *Endocrinology* 70: 638-49.
- Cummings, J. H. 1978. Nutritional implications of dietary fiber. *American Journal of Clinical Nutrition* 31: s21-9.
- DeLuca, H. F. 1979. W. O. Atwater Lecture: The vitamin D system in the regulation of calcium and phosphorus metabolism. *Nutrition Reviews* 37: 161-93.
1980. The vitamin D hormonal system: Implications for bone diseases. *Hospital Practice* 15: 57-66.
1988. The vitamin D story: A collaborative effort of basic science and clinical medicine. *Federation Proceedings* 2: 224-36.
- Draper, H. H., T. Sie, and J. C. Bergan. 1972. Osteoporosis in aging rats induced by high phosphorus diets. *Journal of Nutrition* 102: 1133-42.
- Ebashi, S. 1987. Ca²⁺ in the heart. *Journal of Molecular and Cellular Cardiology* 19: 1-18.
- Feitelberg, S., S. Epstein, F. Ismail, et al. 1987. Deranged bone mineral metabolism in chronic alcoholism. *Metabolism* 36: 322-6.
- Flanagan, B., and G. Nichols. 1969. Bone matrix turnover and balance in vitro. II. The effect of aging. *Journal of Clinical Investigation* 48: 607-12.
- Fogelman, I., L. Smith, R. Mazess, et al. 1986. Absorption of oral diphosphonate in normal subjects. *Clinical Endocrinology* 24: 57-62.
- Foster, G. V., G. F. Joplin, I. MacIntyre, et al. 1966. Effect of thyrocalcitonin in man. *Lancet* 1: 107-9.
- Francis, R. M. 1995. Oral biphosphonates in the treatment of osteoporosis: A review. *Current Therapeutic Research* 56: 831-51.
- Fye, W. B. 1984. Sydney Ringer. Calcium and cardiac function. *Circulation* 69: 849-53.
- Garn, S. M. 1970. Calcium requirements for bone building and skeletal maintenance. *American Journal of Clinical Nutrition* 23: 1149-50.
- Gascon-Barre, M. 1985. Influence of chronic ethanol consumption on the metabolism and action of vitamin D. *Journal of the American College of Nutrition* 4: 565-74.
- Gloth, F. M., and J. D. Tobin. 1995. Vitamin D deficiency in older people. *Journal of American Geriatrics Society* 43: 822-8.
- Goldsmith, R. S., P. Killian, S. H. Ingbar, et al. 1969. Effect of phosphate supplementation during immobilization of normal men. *Metabolism* 18: 349-68.
- Goldsmith, R. S., C. F. Woodhouse, S. H. Ingbar, et al. 1967. Effect of phosphate supplements in patients with fractures. *Lancet* 1: 687-90.
- Greenwald, E., J. Samachson, and H. Spencer. 1963. Effect of lactose on calcium metabolism in man. *Journal of Nutrition* 79: 531-8.
- Hannon, R. R., E. Shorr, W. S. McClellan, and E. F. Du Bois. 1930. A case of osteitis fibrosa cystica (osteomalacia?) with evidence of hyperactivity of the parathyroid bodies. Metabolic study I. *Journal of Clinical Investigation* 8: 215-27.
- Harrison, D. C., and E. Mellanby. 1934. Phytic acid and the rickets-producing action of cereals. *Biochemistry Journal* 33: 1660-80.
- Heaney, R. P. 1982. Calcium intake requirement and bone mass in the elderly. *Journal of Laboratory Clinical Medicine* 100: 309-12.
- Heaney, R. P., J. C. Gallagher, C. C. Johnston, et al. 1982. Calcium nutrition and bone health in the elderly. *American Journal of Clinical Nutrition* 36: 986-1013.
- Heaney, R. P., R. R. Recker, and P. D. Saville. 1977. Calcium balance and calcium requirements in middle-aged women. *American Journal of Clinical Nutrition* 30: 1603-11.
- Hegsted, D. M., I. Moscoso, and C. Collazos. 1952. A study of the minimum calcium requirements of adult men. *Journal of Nutrition* 46: 181-201.
- Hess, A. F., and M. Weinstock. 1924. Antirachitic properties imparted to inert fluids by ultraviolet irradiation. *Journal of the American Medical Association* 83: 1845-6.
- Hirsch, P. F., and P. L. Munson. 1969. Thyrocalcitonin. *Physiological Reviews* 49: 548-622.
- Hodsman, A. B. 1989. Effects of cyclical therapy for osteoporosis using an oral regimen of inorganic phosphate and sodium etidronate: A clinical and bone histomorphometric study. *Bone Mineral* 5: 201-12.
- Hoekstra, W., P. K. Lewis, P. H. Phillips, et al. 1956. The relationship of parakeratosis, supplemental calcium and zinc to the zinc content of certain body components of swine. *Journal of Animal Science* 15: 752-64.
- Huldschinsky, K. 1919. Heilung von Rachitis durch künstliche Höhensonne. *Deutsche Medizinische Wochenschrift* 45: 712-13.
- Irwin, M. I., and E. W. Kienholz. 1973. A conspectus of research on calcium requirements of man. *Journal of Nutrition* 103: 1019-95.
- Ismail-Beigi, F., J. G. Reinhold, B. Faraji, et al. 1977. Effects of cellulose added to diets of low and high fiber content upon the metabolism of calcium, magnesium, zinc and phosphorus by man. *Journal of Nutrition* 107: 510-18.
- Jeans, P., and G. Stearns. 1938. The effect of vitamin D on linear growth in infancy. II. The effects of intakes above 1800 USP units daily. *Journal of Pediatrics* 13: 730-41.
- Kallfelz, F. A., A. N. Taylor, and R. H. Wasserman. 1967. Vitamin D-induced calcium binding factor in rat intestinal mucosa. *Federation Proceedings* 125: 54-8.
- LaFlamme, G. H., and J. Jowsey. 1972. Bone and soft tissue changes with oral phosphate supplements. *Journal of Clinical Investigation* 51: 2834-40.
- Laitinen, K., and M. Valimäki. 1991. Alcohol and bone. *Calcified Tissue International* 49: s70-3.
- Leichsenring, J. M., L. M. Norris, S. A. Lamison, et al. 1951. The effect of level of intake on calcium and phosphorus metabolism in college women. *Journal of Nutrition* 45: 407-18.
- Lotz, M., E. Zisman, and F. C. Bartter. 1968. Evidence for a phosphorus-depletion syndrome in man. *New England Journal of Medicine* 278: 409-15.
- Lowenstein, W. R., and B. Rose. 1978. Calcium in (junctional) intercellular communication and a thought on its

- behavior in intracellular communication. *Annals of the New York Academy of Science* 307: 285-307.
- Luecke, R. W., J. A. Hofer, W. S. Brammell, et al. 1957. Calcium and zinc in parakeratosis of swine. *Journal of Animal Science* 16: 3-11.
- Lukert, B. P., and J. S. Adams. 1976. Calcium and phosphorus homeostasis in man: Effect of corticosteroids. *Archives of Internal Medicine* 136: 1249-53.
- Lumb, G. A., and S. W. Stanbury. 1974. Parathyroid function in human vitamin D deficiency and vitamin D deficiency in primary hyperparathyroidism. *American Journal of Medicine* 56: 833-9.
- Lund, J., and H. F. DeLuca. 1966. Biologically active metabolite of vitamin D₃ from bone, liver, and blood serum. *Journal of Lipid Research* 7: 739-44.
- Lutwak, L., F. R. Singer, and M. R. Urist. 1974. Current concepts of bone metabolism. *Annals of Internal Medicine* 80: 630-44.
- Malm, O. J. 1958. Calcium requirements and adaptation in adult men. *Scandinavian Journal of Clinical Laboratory Investigations* 10 (Supplement 36): 1-290.
- Mandl, F. 1926. Klinisches und Experimentelles zur Frage der lokalisierten und generalisierten Osteitis fibrosa. B. Die generalisierte form der Osteitis fibrosa. *Archiv für Klinische Chirurgie* 143: 1-46.
- Marie, P. J., S. C. Skoryna, R. J. Pivon, et al. 1985. Histomorphometry of bone changes in stable strontium therapy. In *Trace substances in environmental health*, Vol. 19, ed. D. D. Hemphill, 193-208. Columbia, Mo.
- Matkovic, V., K. Kostial, I. Simonovic, et al. 1979. Bone status and fracture rates in two regions of Yugoslavia. *American Journal of Clinical Nutrition* 32: 540-9.
- Mawer, B. E., J. Backhouse, L. F. Hill, et al. 1975. Vitamin D metabolism and parathyroid function in man. *Clinical Science of Molecular Medicine* 48: 349-65.
- Mazess, R. B., H. S. Barden, M. Ettinger, et al. 1988. Bone density of the radius, spine and proximal femur in osteoporosis. *Journal of Bone and Mineral Research* 3: 13-18.
- McCance, R. A., E. M. Widdowson, and H. Lehmann. 1942. The effect of protein intake on the absorption of calcium and magnesium. *Biochemistry Journal* 36: 686-91.
- McCollum, E. V., N. Simmonds, M. Kinney, et al. [1921] 1995. Studies on experimental rickets. XVII. The effects of diets deficient in calcium and in fat-soluble A in modifying the histological structure of the bones. *American Journal of Epidemiology* 141: 280-96.
- McCollum, E. V., N. Simmonds, P. G. Shipley, and E. A. Park. 1921. VIII. The production of rickets by diets low in phosphorus and fat-soluble vitamin A. *Journal of Biological Chemistry* 47: 507-27.
- Mellanby, E. 1921. Experimental rickets. *Medical Research Council Special Report Series* No. 61: 1-78.
1949. The rickets-producing and anticalcifying action of phytate. *Journal of Physiology* 109: 488-533.
- Mendelson, J. H., M. Ogata, and N. K. Mello. 1971. Adrenal function and alcoholism. I. Serum cortisol. *Psychosomatic Medicine* 33: 145-57.
- Mills, R., H. Breiter, E. Kempster, et al. 1940. The influence of lactose on calcium retention in children. *Journal of Nutrition* 20: 467-76.
- Mines, G. R. 1911. On the replacement of calcium in certain neuro-muscular mechanisms by allied substances. *Journal of Physiology* 42: 251-66.
- Moldawer, M., S. J. Zimmerman, and L. C. Collins. 1965. The incidence of osteoporosis in elderly whites and elderly Negroes. *Journal of the American Medical Association* 194: 859-62.
- Morris, E. R., and B. L. O'Dell. 1961. Magnesium deficiency in the guinea pig. Mineral composition of tissues and distribution of acid-soluble phosphorus. *Journal of Nutrition* 75: 77-85.
- National Institutes of Health Consensus Conference. 1995. Optimal calcium intake. *Journal of the American Medical Association* 272: 1942-8.
- National Research Council. 1989. *Recommended dietary allowances*. Tenth edition. Washington, D.C.
- Nordin, B. E. C. 1966. International patterns of osteoporosis. *Clinical Orthopedics and Related Research* 45: 17-30.
- Nordin, B. E. C., M. M. Young, L. Bulusu, et al. 1970. Osteoporosis reexamined. In *Osteoporosis*, ed. U. S. Barzel, 47-67. New York.
- O'Dell, B. L. 1960. Magnesium requirement and its relation to other dietary constituents. *Federation Proceedings* 19: 648-54.
- Olson, E. B., Jr., and H. F. DeLuca. 1973. Vitamin D. Metabolism and mechanism of action. *World Review of Nutrition and Dietetics* 17: 164-88.
- Opie, L. H. 1980. Cardiac metabolism: Catecholamines, calcium, cyclic AMP, and substrates. *Advances in Myocardiology* 1: 3-20.
- Osis, D., L. Kramer, E. Wiatrowski, and H. Spencer. 1974. Dietary fluoride intake in man. *Journal of Nutrition* 104: 1313.
- Paget, J. 1877. On a form of chronic inflammation of bones (osteitis deformans). *Medical Chirurgical Transactions* 60: 37-64.
- Palm, T. A. 1890. Geographical distribution and etiology of rickets. *Practitioner* 45: 270-9.
- Pechet, M. M., E. Bobadilla, E. L. Carroll, et al. 1967. Regulation of bone resorption and formation. *American Journal of Medicine* 43: 696-710.
- Pettifor, J. M., and F. P. Ross. 1983. Low dietary calcium intake and its role in the pathogenesis of rickets. *South African Medical Journal* 63: 179-84.
- Pettifor, J. M., F. P. Ross, R. Travers, et al. 1981. Dietary calcium deficiency: A syndrome associated with bone deformities and elevated 1,25-dihydroxyvitamin D concentrations. *Metabolic Bone Disease and Related Research* 2: 302-95.
- Pettifor, J. M., F. P. Ross, J. Wong, et al. 1978. Rickets in rural blacks in South Africa. Is dietary calcium a factor? *Journal of Pediatrics* 92: 320-4.
- Ponchon, G., and H. F. DeLuca. 1969. Metabolites of vitamin D₃ and their biologic activity. *Journal of Nutrition* 99: 157-67.
- Pramanik, A. K., S. Gupta, and P. S. Agarwal. 1971. Rickets in protein calorie malnutrition. *Indian Pediatrics* 8: 195-9.
- Reinhold, J. G., K. Nasr, A. Lahimgarzaden, et al. 1973. Effects of purified phytate and phytate-rich breads upon metabolism of zinc, calcium, phosphorus and nitrogen in man. *Lancet* 1: 283-8.
- Rich, C., J. Ensink, and P. Ivanovich. 1964. The effects of sodium fluoride on calcium metabolism of subjects with metabolic bone diseases. *Journal of Clinical Investigation* 43: 545-56.
- Ringer, S. 1883. A further contribution regarding the influence of the different constituents of the blood on the contraction of the heart. *Journal of Physiology* 4: 29-42.
1886. Further experiments regarding the influence of small quantities of lime, potassium and other salts on muscular tissue. *Journal of Physiology* 7: 291-308.
- Robertson, I. 1969. A survey of clinical rickets in the infant population in Cape Town in 1967 to 1968. *South African Medical Journal* 43: 1072-6.
- Rubin, R. P. 1982. *Calcium and cellular secretion*. New York.
- Sandler, R. B., C. Slemenda, R. E. LaPorte, et al. 1985. Postmenopausal bone density and milk consumption in

- childhood and adolescence. *American Journal of Clinical Nutrition* 42: 270-4.
- San Fillippo, F. A., and G. C. Battistone. 1971. The fluoride content of a representative diet of the young adult male. *Clinical Chimica Acta* 31: 453-7.
- Schuette, S. A., M. B. Zemel, and H. M. Linkswiler. 1980. Studies on the mechanism of protein-induced hypercalciuria in older men and women. *Journal of Nutrition* 110: 305-15.
- Schwartz, R., N. A. Woodcock, J. D. Blakely, et al. 1973. Metabolic response of adolescent boys to two levels of dietary magnesium and protein. II. Effect of magnesium and protein level on calcium balance. *American Journal of Clinical Nutrition* 26: 519-23.
- Shai, F., R. K. Baker, and S. Wallach. 1971. The clinical and metabolic effects of porcine calcitonin on Paget's disease of bone. *Journal of Clinical Investigation* 50: 1927-40.
- Sherman, H. C. 1920. Calcium requirement of maintenance in man. *Journal of Biological Chemistry* 44: 21-7.
- Sherman, H. C., and A. M. Pappenheimer. 1921. I. A diet producing rickets in white rats and its prevention by the addition of an inorganic salt. *Journal of Experimental Medicine* 34: 189-98.
- Shohl, A. T. 1939. *Mineral metabolism*. American Chemical Society Monograph Series. New York.
- Shorr, E., and A. C. Carter. 1950. The value of strontium as an adjuvant to calcium in the mineralization of the skeleton in osteoporosis in man. In *Metabolic interrelations*, ed. E. C. Reifenshtein, Jr., and V. Johnson, 144-54. New York.
- Spencer, H. 1982. Osteoporosis: Goals of therapy. *Hospital Practice* 17: 131-51.
- Spencer, H., H. Fuller, C. Norris, and D. Williams. 1994. Effect of magnesium on the intestinal absorption of calcium in man. *Journal of the American College of Nutrition* 5: 485-92.
- Spencer, H., A. Hausinger, and D. Laszlo. 1954. The calcium tolerance test in senile osteoporosis. *Journal of the American Geriatric Society* 2: 19-25.
- Spencer, H., and L. B. Kramer. 1985. Factors influencing calcium balance in man. In *Calcium in biological systems*, ed. K. P. Rubin, G. B. Weiss, and J. W. Putney, Jr. New York.
1987. Osteoporosis, calcium requirement and factors causing calcium loss. *Clinics in Geriatric Medicine* 3: 389-402.
- Spencer, H., L. Kramer, M. DeBartolo, et al. 1983. Further studies of the effect of a high protein diet as meat on calcium metabolism. *American Journal of Clinical Nutrition* 37: 924-9.
- Spencer, H., L. Kramer, M. Lesniak, et al. 1984. Calcium requirements in humans. *Clinical Orthopedics and Related Research* 184: 270-80.
- Spencer, H., L. Kramer, C. Norris, et al. 1982. Effect of small doses of aluminum-containing antacids on calcium and phosphorus metabolism. *American Journal of Clinical Nutrition* 36: 32-40.
1984. Effect of calcium and phosphorus on zinc metabolism in man. *American Journal of Clinical Nutrition* 40: 1213-18.
- Spencer, H., L. Kramer, and D. Osis. 1982. Factors contributing to calcium loss in aging. *American Journal of Clinical Nutrition* 36: 776-87.
1988. Do protein and phosphorus cause calcium loss? *Journal of Nutrition* 118: 657-60.
- Spencer, H., L. Kramer, D. Osis, et al. 1978a. Effect of phosphorus on the absorption of calcium and on the calcium balance in man. *Journal of Nutrition* 108: 447-57.
- 1978b. Effect of a high protein (meat) intake on calcium metabolism in man. *American Journal of Clinical Nutrition* 31: 2167-80.
- Spencer, H., L. Kramer, E. Wiatrowski, et al. 1978. Magnesium-fluoride interrelationships in man. II. Effect of magnesium on fluoride metabolism. *American Journal of Physiology* 234: E343-6.
- Spencer, H., M. Lesniak, C. Gatz, et al. 1980. Magnesium absorption and metabolism in patients with chronic renal failure and in patients with normal renal function. *Gastroenterology* 79: 26-34.
- Spencer, H., M. Lesniak, L. Kramer, et al. 1980. Studies of magnesium metabolism in man. In *Magnesium in health and disease*, ed. M. Cantin and M. S. Seelig, 911-19. Jamaica, N.Y.
- Spencer, H., I. Lewin, D. Osis, et al. 1970. Studies of fluoride and calcium metabolism in patients with osteoporosis. *American Journal of Medicine* 49: 814-22.
- Spencer, H., J. Menczel, and I. Lewin. 1964. Metabolic and radioisotope studies in osteoporosis. *Clinical Orthopedics and Related Research* 35: 202-19.
- Spencer, H., J. Menczel, I. Lewin, et al. 1965. Effect of high phosphorus intake on calcium and phosphorus metabolism in man. *Journal of Nutrition* 86: 125-32.
- Spencer, H., C. Norris, and D. Osis. 1992. Further studies of the effect of zinc on intestinal absorption of calcium in man. *Journal of the American College of Nutrition* 5: 561-6.
- Spencer, H., D. Osis, L. Kramer, et al. 1975. Effect of calcium and phosphorus on fluoride metabolism in man. *Journal of Nutrition* 105: 733-40.
- Spencer, H., D. Osis, and M. Lender. 1981. Studies of fluoride metabolism in man. *Science of the Total Environment* 17: 1-12.
- Spencer, H., N. Rubio, M. Indreika, et al. 1986. Chronic alcoholism: Frequently overlooked cause of osteoporosis in men. *American Journal of Medicine* 80: 393-7.
- Spencer, H., N. Rubio, L. Kramer, et al. 1987. Effect of zinc supplements on the intestinal absorption of calcium. *Journal of the American College of Nutrition* 6: 47-51.
- Spencer, H., R. Schwartz, and D. Osis. 1988. Magnesium balances and ²⁸Mg studies in man. In *Trace substances in environmental health*, Vol. 22, ed. D. D. Hemphill, 128-35. Columbia, Mo.
- Stanbury, S. W. 1981. Vitamin D and hyperparathyroidism. *Journal of the Royal College of Physicians* 15: 205-16.
- Stearns, G., P. C. Jeans, and V. Vandecar. 1936. The effect of vitamin D on linear growth in infancy. *Journal of Pediatrics* 9: 1-19.
- Steenbock, H., and A. Black. 1924. XVII. The induction of growth-promoting and calcifying properties in a ration by exposure to ultraviolet light. *Journal of Biological Chemistry* 61: 405-22.
- Strause, L., P. Saltman, K. T. Smith, et al. 1994. Spinal bone loss in postmenopausal women supplemented with calcium and trace minerals. *Journal of Nutrition* 124: 1060-4.
- Theiler, A. 1976. Rickets and other deficiency diseases of the osseous system (last lecture given by Sir Arnold Theiler Onderstepoort, 1936). *Journal of South African Veterinary Association* 47: 41-2.
- Tsai, K. S., H. Heath, R. Kumar, et al. 1984. Impaired vitamin D metabolism with aging in women. Possible role in the pathogenesis of senile osteoporosis. *Journal of Clinical Investigation* 73: 1668-72.
- Wachman, A., and D. S. Bernstein. 1968. Diet and osteoporosis. *Lancet* 1: 958-9.
- Walker, P. M., and H. M. Linkswiler. 1972. Calcium retention in the adult human male as affected by protein intake. *Journal of Nutrition* 102: 1297-1302.

- Warren, J. M., and H. Spencer. 1972. Analysis of stable strontium in biological materials by atomic absorption spectrophotometry. *Clinical Chimica Acta* 38: 435-9.
1976. Metabolic balances of strontium in man. *Clinical Orthopedics and Related Research* 117: 307-20.
- Wasserman, R. H., R. A. Corradina, and A. N. Taylor. 1968. Vitamin D-dependent calcium-binding protein. Purification and some properties. *Journal of Biological Chemistry* 243: 3978-86.
- Wasserman, R. H., and A. N. Taylor. 1968. Vitamin D-dependent calcium-binding protein. Response to some physiological and nutritional variables. *Journal of Biological Chemistry* 243: 3987-93.
- Watts, N. B., S. T. Harris, H. K. Genant, et al. 1990. Intermittent cyclical etidronate treatment of postmenopausal osteoporosis. *New England Journal of Medicine* 323: 73-9.
- Whedon, G. D. 1959. Effects of high calcium intake on bone, blood and soft tissue: Relationship of calcium intake to balance in osteoporosis. *Federation Proceedings* 18: 1112-18.
- WHO (World Health Organization). 1962. *Calcium requirements*. Technical Report Series, No. 230. Geneva.
- Woodhouse, N. J. Y., P. Bordier, M. Fisher, et al. 1971. Human calcitonin in the treatment of Paget's bone disease. *Lancet* 1: 1139-43.
- Woodhouse, N. J. Y., F. H. Doyle, and G. F. Joplin. 1971. Vitamin D deficiency and primary hyperparathyroidism. *Lancet* 2: 283-7.
- Woodhouse, N. J. Y., S. Tun Chot, P. Bordier, et al. 1973. Vitamin D administration in primary hyperparathyroidism. *Clinical Science* 44: 3P.
- Yates, A. J., R. O. C. Oreffo, K. Mayor, et al. 1991. Inhibition of bone resorption by inorganic phosphate is mediated by both reduced osteoclast formation and decreased activity of mature osteoclasts. *Journal of Bone and Mineral Research* 6: 473-8.
- Zipkin, I., A. S. Posner, and E. D. Eanes. 1962. The effect of fluoride on the X-ray-diffraction pattern of the apatite of human bone. *Biochimica Biophysica Acta* 59: 255-8.

IV.B.2 Iodine and Iodine-Deficiency Disorders

The term "iodine-deficiency disorders" (IDD) is now used to denote all the effects of iodine deficiency on growth and development (Hetzel 1983). In the past, the term "goiter" was used to describe such effects, but IDD has now been generally adopted in the field of international nutrition and health. In the last 10 years, this reconceptualization has helped to focus more attention on the problem of iodine deficiency. For much of our historical treatment, however, we use the terms "goiter" and "IDD" interchangeably.

Extensive reviews of the global geographic prevalence of goiter have been published. One of these, by F. C. Kelly and W. W. Snedden, appeared as a World Health Organization (WHO) monograph in 1960. A second survey was done more recently by J. Stanbury and B. Hetzel (1980), and the reader is referred to these sources for a closer look at the many countries

that still have a significant goiter problem. In general, goiter is associated with elevated areas and regions where there has been leaching of iodine from the soil due to glaciation, snow water, high rainfall, or floods. Thus the great mountain chains of the world, the European Alps, the Himalayas, and the Andes, have become well known as goiter areas, and in fact most mountainous regions throughout the world have iodine-deficient areas.

We now know that in addition to mountainous areas, flooded river valleys, such as those of the Ganges, Brahmaputra, and Irawaddy Rivers in Southeast Asia, have their soils leached of iodine and thus are also severely deficient in the mineral. These inescapable geographical facts mean that vast populations are at risk of iodine-deficiency disorders, and unfortunately it is likely that soil erosion in modern times is acting to increase the iodine-deficient areas of the world. It is, of course, axiomatic that populations totally dependent on food grown in such soil, as in systems of subsistence agriculture, will become iodine deficient.

History of Goiter and Cretinism

Descriptions and speculations about goiter and cretinism (the best-known iodine-deficiency disorders) go back to the ancient world. A historical review provides a fascinating succession of cultural concepts culminating in the twentieth century, when the causative role of iodine deficiency was established and the control of the disorders demonstrated. In the waning years of the twentieth century a global action program was organized in order to eliminate this ancient scourge of humankind by the beginning of the new millennium.

The Ancient Civilizations

An old reference to goiter is attributed to a mythical Chinese cultural hero, Shen-Nung. In his book *Shen-nung pen-ts'ao ching* (the divine husbandman's classic on materia medica), he is said to mention the seaweed sargassa as an effective remedy for goiter. Goiter is also mentioned in the *Shan Hai Ching* (which probably reached its present form about the second century B.C.), which attributes the disease to the poor quality of water. Other references during the Han dynasty (206 B.C. to A.D. 220) and the Wei dynasty (A.D. 200-264) mention deep mental emotions and "certain conditions of life in the mountain regions" as causes of goiter. The treatment of goiter with sargassa weeds is mentioned by the famous early-fourth-century Chinese medical writer Ge Hong. The Chinese also employed animal thyroid in the treatment of goiter - the use of deer thyroid is mentioned in a sixth-century text. Animal thyroid continued to be used in China and is discussed again by an eminent Chinese physician, Li Shih-chen, in his well-known 1596 herbal *Pen-ts'ao kang-mu* (materia medica arranged according to drug descriptions and technical aspects), in

which preparations of pig and deer thyroid are mentioned. The continued use of seaweed and animal thyroid over so many hundreds of years in China suggests that there was certainly some benefit derived from these measures in the treatment of goiter.

Elsewhere in the ancient world there is less mention of the disease. In ancient Hindu literature incantations against goiter may be found in the Veda *Attharva* dating from around 2000 B.C. According to the Ebers papyrus, tumors of the neck were also known in ancient Egypt, where they were treated surgically.

In the famous fourth-century B.C. volume *Airs, Waters, and Places*, attributed to Hippocrates, drinking water was regarded as a cause of goiter. At about the beginning of the Christian era, Aulus Celsus described a fleshy tumor of the neck, which he incised to find that it contained honey-like substances and even small bones and hairs (probably a long-standing goiter). It was subsequently deduced by the Roman physician Galen in the second century A.D. that the glands of the neck, including the thyroid, had the function of secreting a fluid into the larynx and the pharynx. Such views continued to be accepted for many centuries and were held by such famous seventeenth- and eighteenth-century physicians as Marcello Malpighi and Herman Boerhaave.

Medieval Europe

It is of interest to note the attention given to goiter in paintings and sculptures of the Middle Ages. The late Professor F. Merke of Berne unearthed a number of fascinating examples from manuscripts and churches. The earliest depiction of goiter and cretinism is in a book in the Austrian National Library in Vienna – the *Reuner Musterbuch* – dating from 1215. The book was originally from the Cistercian Abbey in Reun near the city of Graz in Styria (Austria), where goiter was highly endemic until recent times. The picture (Figure IV.B.2.1) shows a figure with three large goiters and a stupid facial expression, brandishing a fool's staff in one hand while reaching up with the other toward a toad. This was doubtless a goitrous cretin. It was common to depict a fool grasping a cudgel as shown in this figure. The significance of the toad may be related to the popular use of live or dismembered frogs for the treatment of goiter. As Merke points out, this picture of the Reun cretin predates by some 300 years the recognition of the relationship between goiter and cretinism by Paracelsus (Philippus Aureolus Theophrastus Bombastus von Hohenheim).

According to F. de Quervain and C. Wegelin (1936), the term “cretin” most likely comes from the words “Christianus” or “Crestin” in the southeastern French dialect – referring to a *bon chrétien* because of the innocence of these subjects. These sufferers were in fact given special recognition in the medieval world and often regarded as angels or innocents with magical powers (Merke 1984). Clearly the Church did its best to fit them into the prevailing religious culture.

The Renaissance

Felix Plattner of Basel wrote a classic description of goiter and cretinism following a 1562 visit to the Valais:

In Bremis, a village of the Valais, as I have seen myself, and in the Valley of Carinthia called Binstzgerthal [today the Pinzgau], it is usual that many infants suffer from innate folly [simple-mindedness]. Besides, the head is sometimes misshapen: the tongue is huge and swollen; they are dumb; the throat is often goiterous. Thus, they present an ugly sight; and sitting in the streets and looking into the sun, and putting little sticks in between their fingers [a stick resting between their hands], twisting their bodies in various ways, with their mouths agape they provoke passersby to laughter and astonishment. (Langer 1960: 13–14)



Figure IV.B.2.1. “The Reun cretin” from the Reun Model Book, *Reuner Musterbuch*, produced by the Cistercian Abbey at Reun (near Graz), Austria, and dating from 1215. This model book covers everything that an educated “clericus” of the thirteenth century had to know. The cretin has a large trilobed goiter with the third lobe strung over the left shoulder. His right hand clutches a frog (possibly for healing), and his left hand holds the “fool’s scepter” in order to make the mental deficiency clear. (From Merke 1984, with permission.)



Figure IV.B.2.2. Madonna and child by Francesco di Gentili, fifteenth century. The madonna has a goiter and her child looks hypothyroid. (Pinacoteca, Vatican Museums, Rome; photo, Art Resource, New York.)

A study of goiter in sixteenth-century art was made by H. Hunziger (1915); the disease can be readily observed in the Madonnas of the Renaissance. Works by masters such as Jan van Eyck and Lucas Van der Leyden portray it, as do other paintings in the Sienese and Vatican galleries (Figure IV.B.2.2). All this may indicate that the condition was virtually accepted as normal because it was so common. In a later period goiter is also to be seen in the paintings of Peter Paul Rubens, Rogier van der Weyden, and Albrecht Dürer.

During the Renaissance, goiter was believed to be curable by the touch of a king. According to the king's personal physician, Henry IV of France caused 1,500 goiters to regress by touching the patient and using the formula "Le Roi te touche et Dieu te guerit." The "touch" was also practiced by many English kings; Charles II is alleged to have touched 9,200 persons suffering from the "King's Evil" or scrofula (with which goiter was often confused). According to newspaper reports, on March 20, 1710, Queen Anne revived the ancient custom of curing goiter and scrofula by the laying on of hands. As we pointed out, however, in the sixteenth century Paracelsus had rec-

ognized the association of goiter and cretinism and attributed the disease to a deficiency of minerals in drinking water. This reasoning rather neatly contrasts the scientific approach with the magic of the King's Touch!

The Seventeenth Century

The study of human anatomy prospered following the sixteenth-century pioneering work of Andreas Vesalius, Matteo Realdo Colombo, Hieronymus Fabricius ab Aquapendente, and Bartolomeo Eustachi. All of these anatomists noted the thyroid gland, which was called the "glandulus laryngis" by Vesalius. Fabricius recognized the connection between the glandulus laryngis and goiter. But the first clear description of the gland was penned in Latin in 1656 by an Englishman, Thomas Wharton (the translated title is *Adenography, or a Description of the Glands of the Entire Body*). Nonetheless, in the seventeenth century the function of the thyroid was not understood. It was usually regarded as secreting a fluid to "humidify" the walls of the larynx, the pharynx, and the trachea. In fact, the function of the thyroid was not understood until the latter part of the nineteenth century, when it was recognized to have an "internal" secretion in the form of the thyroid hormone, and not an external one.

The Eighteenth Century

In the eighteenth century there was a great escalation in scientific observations and the reporting of such observations. Many of these were collected in Denis Diderot's *Encyclopédie* (1751-72), in which the term "cretin" appeared in print for the first time in an article by Diderot's co-editor Jean Le Rond d'Alembert. His definition of a cretin was that of "an imbecile who is deaf, dumb with a goiter hanging down to the waist."

The Nineteenth Century

Interest in and concern about the possibility of the control of goiter accelerated in the early nineteenth century when Napoleon Bonaparte ordered a systematic investigation of the disease. He did so because large numbers of young men from certain regions were being rejected as unfit for military duties. Moreover, Napoleon himself had probably seen something of the problem during his march into Italy through the goiter-infested Valais.

Iodine was isolated from the ashes of the seaweed *Fucus vesicularis* by B. Courtois in France in 1811, and in 1820 Jean François Coindet recommended iodine preparations for the treatment of goiter. However, soon afterward marked opposition developed to its employment because of the occurrence of symptoms of toxicity, which we now know was the result of excessive thyroid secretion.

The iodination of salt was first suggested by Jean Baptiste Boussingault, who resided for many years in Colombia in South America. The people among whom he lived obtained their salt from an abandoned mine

and felt that this salt conferred special health benefits. In 1825 Boussingault analyzed the salt and found that it contained large quantities of iodine. In 1833 he suggested that iodized salt be used for the prevention of goiter. Unfortunately, an experiment carried out in France once more obscured the importance of iodine in the etiology of the disease. Goitrous families received salt fortified with 0.1 to 0.5 grams (g) of potassium iodide per kilogram (kg) of salt. But the high dosage produced symptoms of an excess thyroid secretion and consequently the treatment again fell into disrepute.

The Twentieth Century

Present-day practice in the prevention and control of goiter is based on the work of David Marine, who in 1915 declared that “endemic goitre is the easiest known disease to prevent.” Marine and his colleague, O. Kimball, carried out the first large-scale trials with iodine in Akron, Ohio, from 1916 to 1920. About 4,500 girls between 11 and 18 years of age took part in the experiment. Roughly half of this group had goiter; the other half had normal thyroid activity. Of the group, 2,190 girls were given a daily dose of 0.2 g of sodium iodide in water for 10 days in the spring and 10 days in the autumn, making a total dose of 4.0 g over the year. The remaining 2,305 girls acted as controls. Two facts stand out from the data generated by this experiment: (1) In the group receiving sodium iodide, of 908 girls with a normal thyroid prior to treatment, only 2 (0.2 percent) developed goiter; however, in the control group, of 1,257 girls with previously normal thyroid, goiter appeared in 347 (27.6 percent); (2) in the group treated for goiter, 773 out of 1,282 girls with the disease (60.4 percent) showed a considerable decrease in the size of the thyroid, whereas in the control group, spontaneous regression of the goiter occurred in only 145 out of 1,048 girls (13.8 percent).

Thus both the prophylactic and the therapeutic effects were impressive. Iodism was very rare (only 11 cases) in spite of the extremely large doses of iodine, and the symptoms disappeared within a few days of stopping the administration of sodium iodide.

Mass prophylaxis of goiter with iodized salt was first introduced in 1924 on a community scale in Michigan (Kimball 1937), where it seems probable that the last glaciation had rendered the soil iodine deficient in that state and throughout much of the Great Lakes region. Goiter surveys of schoolchildren and iodine analyses of their drinking water were carried out in four representative counties, where the average goiter rate among 65,537 children was 38.6 percent. Table salt containing 1 part in 5,000 of potassium iodide was then introduced into the children's diet. By 1929 the average goiter rate had fallen to 9 percent. Moreover a follow-up survey conducted by B. Brush and J. Altland (1952) on 53,785 schoolchildren in the same counties showed a goiter rate of only 1.4 percent. It was also reported that in seven large

hospitals in Michigan thyroidectomies accounted for only 1 percent of all operations in 1950 compared with 3.2 percent in 1939. No toxic symptoms of iodide prophylaxis were observed.

The impact of iodized salt on the control of goiter was also vividly demonstrated in Switzerland. As this country is situated in the elevated region of the European Alps, the burden of goiter and cretinism was great throughout the country. In 1923, for example, the Canton of Berne, with a population of about 700,000, had to hospitalize 700 cretins incapable of social life. But with the Cantons' introduction of iodized salt, which proceeded throughout the 1920s, goiter rates fell steeply, and “deaf and dumb institutions” were later closed or used for other purposes.

Observations of Swiss Army recruits revealed definite evidence of this trend of rapid decline throughout the country. Between the years 1925 and 1947, the number of exemptions for military service fell from 31 to less than 1 per thousand. Moreover, following 60 years of the use of iodized salt in Switzerland, a recent review has made clear the benefits it provided in the prevention of all degrees of neurological damage in the Swiss population (Burgi, Supersaxo, and Selz 1990). Indeed, it is now clear that the “spontaneous” disappearance of cretinism throughout Europe was due to a dietary increase in iodine intake. The continued persistence of iodine deficiency and, consequently, goiter and cretinism in Europe is mainly associated with more isolated rural areas that have not undergone the sort of social and economic development that leads to dietary diversification. But even these cases could be completely prevented with the effective distribution of iodized salt.

Goiter and Iodine Deficiency

Iodine deficiency causes depletion of thyroid iodine stores with reduced daily production of the thyroid hormone (T_4). A fall in the blood level of T_4 triggers the secretion of increased amounts of pituitary thyroid stimulating hormone, which increases thyroid activity with hyperplasia of the thyroid. An increased efficiency of the thyroid iodide pump occurs with faster turnover of thyroid iodine, which can be seen in an increased thyroidal uptake of radioactive isotopes ^{131}I and ^{125}I . These features were first demonstrated in the field in the now classic observations of Stanbury and colleagues (1954) in the Andes of Argentina.

Iodine deficiency is revealed by a determination of urine iodine excretion using either 24-hour samples or, more conveniently, casual samples with determination of iodine content per gram of creatinine. Normal iodine intake is 100 to 150 micrograms per day ($\mu\text{g}/\text{day}$), which corresponds to a urinary iodine excretion in this range (Stanbury and Hetzel 1980). In general, in endemic goiter areas the intake is well below 100 $\mu\text{g}/\text{day}$ and goiter is usually seen when the

level is below 50 µg/day (Pretell et al. 1972). The rate increases as the iodine excretion falls so that goiter may be almost universal at levels below 10 µg/day. The iodine content of drinking water is also low in areas with endemic goiter (Karmarkar et al. 1974).

Goiter, however, can also arise from causes other than iodine deficiency. Foremost among these are a variety of agents known as goitrogens. Recent research (Ermans et al. 1980; Delange, Iteke, and Ermans 1982) has shown that staple foods from the developing world such as cassava, maize, bamboo shoots, sweet potatoes, lima beans, and millets contain cyanogenic glucosides that are capable of liberating large quantities of cyanide by hydrolysis. Not only is the cyanide toxic but the metabolite in the body is predominantly thiocyanate, which is a goitrogen. Fortunately, these glycosides are usually located in the inedible portions of most plants, or if in the edible portion, in such small quantities that they do not cause a major problem. But such is not the case with cassava, which is cultivated extensively in developing countries and represents an essential source of calories for tropical populations of more than 200 million (Delange et al. 1982).

The role of cassava in the etiology of endemic goiter and endemic cretinism was demonstrated by F. Delange and colleagues (1982) in studies conducted in nonmountainous Zaire, and their observations were confirmed by G. Maberly and colleagues (1981) in Sarawak, Malaysia. One major effect of cassava consumption can be to increase iodine loss from the body by increasing urinary excretion.

It is important to note, however, that chronic consumption of large quantities of cassava does not necessarily result in the development of endemic goiter. Such development depends on the balance between the dietary supply of iodine and the thiocyanate (SCN) generated from the cyanide (hydrocyanic acid [HCN]) content of the cassava. The cyanide content of cassava varies with the linamarin content, and this in turn varies with traditional detoxification processes – including soaking in water before consumption, which greatly reduces the HCN content. However, sun-drying of cassava (as in the Ubangi region of Zaire) is not effective in reducing the HCN content, which can, in turn, produce a high prevalence of goiter and endemic cretinism.

A normal or high iodine intake will protect against the goitrogenic effect of the SCN. In fact, a Belgian group has shown that an iodine to SCN (I:SCN) ratio greater than 7 will achieve this protection. Goiter occurs if the ratio is about 3 and will be of high prevalence when the ratio is below 2. At this low ratio endemic cretinism will also be found in the population.

In summary, four factors will determine the SCN ratio: (1) the level of iodine intake in the diet, (2) the HCN content of fresh cassava roots and leaves, (3) the efficiency of the detoxification process used during

the preparation of cassava-based foods, and (4) the frequency and quantity of consumption of these foods.

Cretinism and Iodine Deficiency

The gradual disappearance of cretinism in Europe in the early decades of the twentieth century led to the condition being largely forgotten. But in the 1960s endemic cretinism was virtually rediscovered almost simultaneously in a number of the more remote areas of the world, among them New Guinea, Zaire, India, Indonesia, China, and Brazil.

The clinical manifestations of the condition found in these areas have now been reported in considerable detail (Pharoah et al. 1980), and field studies of thyroid gland function and iodine metabolism have also been made in many countries. Evidence of the association of cretinism with severe iodine deficiency and high goiter rates has been uniformly reported.

One noticeable feature of the condition since its rediscovery has been the occurrence of a wide range of defects in individuals. These range from isolated mental deficiency to deaf-mutism of varying degree to a varying severity of paralysis of the arms and legs. But there are also individuals who appear to be normal apart from some coordination defect, all of which indicates that endemic cretinism is part of a spectrum of defects within an iodine-deficient population. And endemic cretinism is, as we shall see, a community or population disease.

It was R. McCarrison, reporting in 1908 from what was then called the North West Frontier of India (including the Karakoram Mountains, which are now in northern Pakistan), who first clearly distinguished two types of endemic cretins – the “nervous” and the “myxedematous” – from a series of 203 patients whom he had studied. In the nervous type, he recognized mental defect, deaf-mutism, and a spastic diplegia (paralysis) with a spastic rigidity, affecting the legs predominantly, that produced a characteristic walk or gait. Squint was also noted. By contrast, the myxedematous type had all the characteristics of severe hypothyroidism: dry swollen skin and tongue, deep hoarse voice, apathy, and mental deficiency. McCarrison regarded this condition as identical with “sporadic cretinism.” He found deaf-mutism present in 87 percent of his 203 cases.

These two types of cretinism are also readily seen in China, where the neurological type is the predominant form. In Hetian district in Xinjiang (only some 300 kilometers east of Gilgit, where McCarrison made his original observations in the first decade of the twentieth century), a similar pattern was observed in 1982 (Figure IV.B.2.3) with neurological, hypothyroid, and mixed types present (Hetzl 1989). Later, these observations in China were confirmed by S. Boyages and colleagues (1988).



Figure IV.B.2.3. A dwarfed cretin from Xingjiang, China, who is also deaf-mute. This condition is completely preventable. Right: the “barefoot doctor” from her village. Both women are about 35 years of age. (Courtesy of Dr. T. Ma, Tianjin, China.)

It is apparent that these two types of cretinism are distinct conditions. As McCarrison recognized, the features of the myxedematous type are essentially the same as those of sporadic cretinism. However, the myxedematous type is associated with the occurrence of hypothyroidism, which causes endemic cretinism, the latter a result of endemic goiter triggered by severe iodine deficiency in an entire community. By contrast, sporadic cretinism occurs all over the world whether or not iodine deficiency is present. It is usually found with evidence of an absent or misplaced thyroid or with a congenital defect in the biosynthesis of the hormone. To prevent confusion, the term “congenital hypothyroidism” is now generally preferred to sporadic cretinism.

It is now established that endemic cretinism, in its fully developed form characterized by severe mental deficiency, deaf-mutism, and spastic diplegia, is epi-

demologically associated with high rates of goiter and severe iodine deficiency (levels of 25 $\mu\text{g/day}$ or less compared to a normal intake of 80–150 $\mu\text{g/day}$), whereas goiter alone is seen at levels below 50 $\mu\text{g/day}$ (Pharoah et al. 1980; Hetzel and Maberly 1986; Hetzel, Potter, and Dulberg 1990). However, as mentioned previously, the apparent spontaneous disappearance of endemic cretinism in southern Europe raised considerable doubt about its relationship to iodine deficiency (Hetzel 1989).

The hypothesis that there was such a relationship was tested for the first time in a controlled trial beginning in the 1960s. Such testing was made possible by the development of a new method for correction of severe iodine deficiency in Papua New Guinea in the form of the injection of iodized oil (McCullagh 1963; Buttfield et al. 1965; Buttfield and Hetzel 1967). Iodized oil or saline injections were given to alternate families in the Jimi River District in the Western Highlands at the time of the first census (1966). Each child born subsequently was examined for evidence of motor retardation, as assessed by the usual measurements involving sitting, standing, or walking, and for evidence of deafness. Examination was carried out without knowledge as to whether the mother had received iodized oil or saline injections (Pharoah, Buttfield, and Hetzel 1971).

Infants presenting with a full syndrome of hearing and speech abnormalities, together with abnormalities of motor development with or without squint, were classified as suffering from endemic cretinism. By these criteria, there were 7 cretins born to women who had received iodized oil out of a total of 687 children. In 6 of these 7 cases, conception had occurred prior to the iodized oil injections (Figure IV.B.2.4). In the untreated group, there were 25 endemic cretins out of a total of 688 children born since the trial began. In 5 of these 25, conception had occurred prior to saline being given.

It was concluded that an injection of iodized oil given prior to a woman's pregnancy could prevent the occurrence of the neurological syndrome of endemic cretinism in the infant. The occurrence of the syndrome in those who were pregnant at the time of oil injection indicated that the damage probably occurred during the first half of pregnancy (Pharoah et al. 1971).

Subsequent studies in Papua New Guinea have revealed a motor coordination defect in apparently normal children subjected to severe iodine deficiency during pregnancy (Connolly, Pharoah, and Hetzel 1979; Pharoah et al. 1981). There is also evidence of an associated intellectual defect (Pharoah et al. 1984; Fierro-Benitez et al. 1986). Thus it is apparent that the effects of severe iodine deficiency in pregnancy go beyond endemic cretinism to affect children in other unpleasant ways. Such observations have been confirmed in Indonesia and Spain (Bleichrodt et al. 1987) and in China (Boyages et al. 1989).

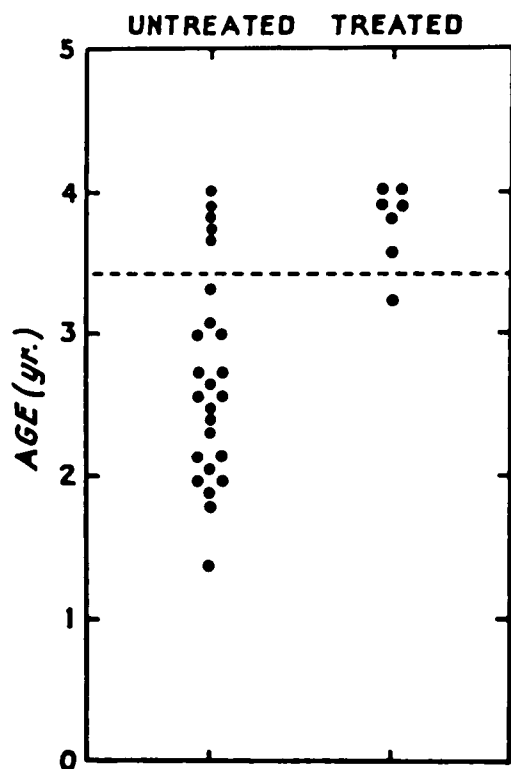


Figure IV.B.2.4. The results of a controlled trial of iodized oil injection in the Jimi River district of the highlands of Papua New Guinea. Alternate mothers were given an injection of iodized oil and saline in September 1966. All newborn children were followed up for the next five years. Each dot represents a cretin child. The figure shows that mothers given iodized oil injections do not have subsequent cretin children, in comparison with the persistence in the untreated group. (From Pharaoh et al. 1971, with permission.)

The importance of the link between iodine deficiency and brain development has led to animal model studies that shed light on the relationship and the mechanisms involved. These studies have revealed reduced brain weight with a reduced number of brain cells during and at the end of pregnancy in iodine-deficient rats, sheep, and marmosets (Hetzel, Chevadev, and Potter 1988). The effect of iodine deficiency is mediated through reduced secretion (by both the maternal and fetal thyroids), which in turn is associated with severe fetal hypothyroidism. To date, the syndrome of neurological cretinism has not been reproduced in an animal species (Hetzel et al. 1988).

The Iodine-Deficiency Disorders

Our concept of the effects of iodine deficiency has undergone a rapid evolution since 1970. Originally

the problem was designated as goiter, which (although a fascinating phenomenon for many thyroidologists) cannot by itself justify a high priority for prevention programs in developing countries. As noted in the introduction, the various effects of iodine deficiency at different stages of life are now included in IDD, which has been generally adopted (Hetzel 1983; *Lancet* 1983). These various disorders (listed in Table IV.B.2.1) occur in populations subjected to iodine deficiency, and all can be prevented by correction of the deficiency. We will now review IDD in detail by reference to the four stages of life in which it occurs.

Iodine Deficiency in the Fetus

In iodine-deficient areas there is an increased rate of abortions and stillbirths, which can be reduced by correction of iodine deficiency (McMichael, Potter, and Hetzel 1980). An increased rate of stillbirths has also been observed in iodine-deficient sheep. In iodine-deficient pastures, lamb losses can be reduced by correction of iodine deficiency, and, it has been suggested, the same is true for goats (Hetzel and Maberly 1986). Pregnancy losses have also been produced experimentally with hypothyroidism. Studies in hypothyroid guinea pigs (produced by surgical removal of the thyroid) reveal a three- to fourfold increase in abortions and stillbirths that can be virtually eliminated by replacement therapy with thyroxine during pregnancy (McMichael et al. 1980).

Table IV.B.2.1. *The spectrum of iodine-deficiency disorders (IDD)*

Fetus	Abortions Stillbirths Congenital anomalies Increased perinatal mortality Increased infant mortality Neurological cretinism mental deficiency deaf-mutism spastic diplegia squint
Neonate	Myxedematous cretinism dwarfism mental deficiency psychomotor defects
Child and adolescent	Neonatal goiter Neonatal hypothyroidism Goiter Juvenile hypothyroidism Impaired mental function Retarded physical development
Adult	Goiter with its complications Hypothyroidism Impaired mental function Iodine-induced hyperthyroidism

In this same vein, data from Zaire and Papua New Guinea indicate an increased perinatal mortality that reflects an increase in stillbirths. There is also increased infant mortality. In Zaire, the results of a controlled trial of iodized oil injections given in the latter half of pregnancy revealed a substantial fall in perinatal and infant mortality with improved birth weight (Thilly 1981). Low birth weight is generally (whatever the cause) associated with a higher rate of congenital anomalies.

Recent evidence indicates that the various effects of iodine deficiency on the fetus (including abortion, stillbirth, congenital anomalies, and the varying manifestations of cretinism) probably arise from the lowered level of thyroid hormone T_4 in the iodine-deficient mother. The more severe the reduction in the level of maternal T_4 , the greater the threat to the integrity of the fetus – a proposition supported by animal data on the causes of abortions and stillbirths (Hetzel et al. 1990).

Iodine Deficiency in the Neonate

In former times neonatal goiter was commonly seen in iodine-deficient areas (Hetzel and Maberly 1986). Neonatal hypothyroidism remains a well-recognized cause of mental defect in the Western world. This is because the development of the brain is dependent on an adequate supply of thyroxine. Only about one-third of normal brain development occurs before delivery of the infant; the other two-thirds is completed in the first two years of life. Hence a normal level of thyroxine is extremely important both during and after pregnancy.

In many countries every newborn child is tested for the level of blood thyroxine (usually taken from a prick in the heel at the fourth and fifth days of life). If, after further investigation, the level is found to be low, and a lowered state of thyroid gland function is indicated, then replacement treatment with daily tablets of thyroxine is begun immediately. The results of such a procedure have been evaluated in several countries and deemed to be generally excellent, provided treatment is given early, meaning within the first month of life. In Western countries, the incidence of such an abnormality runs at 1 per 3,500 live births, and results from (1) the absence of the thyroid; (2) an abnormal position of the thyroid (ectopia); or (3) a defect in the “biochemical machinery” required to produce thyroxine (called a biosynthetic defect) (Burrow 1980).

A much higher incidence, however, is found in iodine-deficient environments. Indeed, observations of blood taken from the umbilical vein (in the umbilical cord), just after birth, have revealed a rate of neonatal hypothyroidism of some 10 percent in Zaire and 5 to 10 percent in northern India and Nepal (Kochupillai and Pandav 1987). In the most severely iodine-deficient environments (more than 50 percent have urinary iodine below 25 $\mu\text{g}/\text{gram}$ creatinine, which is 25 percent of normal intake), the incidence

of neonatal hypothyroidism is 75 to 115 per 1,000 births. By contrast in Delhi, where only mild iodine deficiency is present with a low prevalence of goiter and no cretinism, the incidence drops to 6 per 1,000. But in control areas without goiter the level is only 1 per 1,000.

In India chemical observations have been extended to studies of the Intelligence Quotient (IQ) and hearing ability of schoolchildren. It has been shown that there is a marked reduction in IQ scores in villages with a high rate of (chemical) hypothyroidism in comparison with villages without iodine deficiency. Nerve deafness is also much more common, and there is evidence of lowered growth hormone levels in the blood.

The blood level of thyroxine in the neonate is an important indicator of the severity of iodine deficiency. Monitoring of these levels is now being done in India on samples of cord blood taken by birth attendants and subsequently spread on specially prepared filter paper. The samples are then mailed to a biochemistry laboratory where radioimmunoassay is carried out. This arrangement offers a very convenient and inexpensive method for the assessment of the severity of iodine deficiency provided a suitable functioning laboratory is available. The equipment is complex and costly but automated so that large numbers of samples can be processed. Such laboratories are now being gradually established in India, China, and Indonesia, where there are very large iodine-deficient populations. Data from these laboratories indicate a much greater risk of mental defect in severely iodine-deficient populations than is indicated by the presence of cretinism. Unfortunately, because of a gross inadequacy of manpower and money, it is just not possible to diagnose and treat these infants in most developing-world countries as they are treated in the West. This problem, of course, points up the overwhelming importance of prevention.

Iodine Deficiency in Infancy and Childhood

Iodine deficiency in children is characteristically associated with goiter, the classification of which has now been standardized by the World Health Organization (Dunn et al. 1986). Girls have a higher prevalence than boys and the rate increases with age so that it reaches a maximum with adolescence. Thus observations of goiter rates in schoolchildren between the ages of 8 and 14 provide a convenient indication of the presence of iodine deficiency in a community. The congregation of children in schools is a great advantage in providing access to a population and a place where collection of urine samples for detection of urinary iodine can be conveniently carried out.

As already noted, the prevention and control of goiter in schoolchildren was first demonstrated by Marine and Kimball in Akron, Ohio, in 1917. However, Marine and Kimball (1921) recognized that goiter was only part of IDD. They said:

The prevention of goitre means vastly more than cervical (neck) deformities. It means in addition, the prevention of those forms of physical and mental degeneration such as cretinism, mutism and idiocy which are dependent on thyroid insufficiency. Further it would prevent the development of thyroid adenomas which are an integral and essential part of endemic goitre in man and due to the same stimulus. (Marine and Kimball 1921: 1070)

Recent studies of schoolchildren living in iodine-deficient areas in a number of countries indicate impaired school performance and lower IQ scores in comparison with matched groups from non-iodine-deficient areas. In Papua New Guinea, for example, children involved in the controlled trial in the Western Highlands have been tested periodically. Differences in bimanual dexterity were revealed by threading beads and putting pegs into a pegboard – differences that correlated with whether or not the mother was severely iodine deficient (Connolly et al. 1979). Indeed, critical studies using a wide range of psychological tests have shown that the mental development of children from iodine-deficient areas lags behind that of children from non-iodine-deficient areas. But the differences in psychomotor development became apparent only after the age of 2½ years (Bleichrodt et al. 1987; Boyages et al. 1989).

Another question has been whether these differences can be affected by correction of the iodine deficiency. In a pioneering study initiated in Ecuador by R. Fierro-Benitez and colleagues (1986), the long-term effects of iodized oil injections have been assessed by comparison of two highland villages. In one of them (Tocachi) such injections were administered. The other (La Esperanza) was used as a control. Particular attention was paid to 128 children aged 8 to 15 whose mothers had received iodized oil prior to the second trimester of pregnancy and, of course, a matched control group of 293 children of similar age. All children were periodically examined from birth at key stages in their development. Women in Tocachi were injected or reinjected in 1970, 1974, and 1978. Assessments in 1973, 1978, and 1981 revealed the following:

Scholastic achievement was better in the children of treated mothers when measured in terms of school year reached, for age, school dropout rate, failure rate, years repeated and school marks. There was no difference between the two groups by a certain test (Terman-Merrill, Wechsler scale, or Goodenough). Both groups were impaired in school performance – in reading, writing and mathematics, but more notably the children of untreated mothers. (Fierro-Benitez et al. 1986: 195)

The results indicate the significant role of iodine deficiency, but other factors were also important in the

school performance of these Ecuadorean children, such as social deprivation and other nutritional difficulties.

In 1982 the results were reported of a controlled trial carried out with oral iodized oil in the small Bolivian Highland village of Tiquipaya, 2,645 meters above sea level (Bautista et al. 1982). Each child in the treatment group received an oral dose of 1.0 milliliter (ml) iodized oil (475 mg iodine), whereas children in the second group received 1.0 ml of a mineral oil containing no iodine but of a similar brown color. Subsequent assessment was double-blind so that the group of the child was not known to the examiners. Interestingly, on follow-up 22 months later, the urinary iodine had increased and goiter size had decreased in both groups. This reflected “contamination” of the village environment with iodine, probably the result of fecal excretion by those who had received the iodized oil injections. There were no differences between the two treatment groups in growth rate or in the performance with the tests. However, improvement in IQ could be demonstrated in all those children, regardless of their group, who showed significant reduction of goiter. This was particularly so in girls. It was concluded that correction of iodine deficiency may improve the mental performance of school-age children, but that a bigger dose should be given.

Such studies are now being conducted in a number of countries. More data are required. However, data generated to date point to significant improvement in the mental performance of schoolchildren by the correction of iodine deficiency.

Iodine Deficiency in the Adult

The common effect of iodine deficiency in adults is goiter, although characteristically there is an absence of classical clinical hypothyroidism in adults with endemic goiter. Nonetheless, laboratory evidence of hypothyroidism with reduced T_4 levels is common, associated with apathy, lethargy, and lack of energy.

Iodine administration in the form of iodized salt, iodized bread, or iodized oil have all been demonstrated to be effective in the prevention of goiter in adults (Hetzel 1989). Iodine administration may also reduce existing goiter, particularly in the case of iodized oil injections. The clearly observed benefits of iodine supplementation promote ready acceptance of such measures by people living in iodine-deficient communities.

Such benefits have been well described by J. Li and X. Wang (1987) for the northeast Chinese village of Jixian. In 1978 Jixian had a population of 1,313 people, of whom 65 percent had goiter and 11.4 percent were cretins. The latter included many severe cases and the village was known locally as “the village of idiots.” The economic development of the village was retarded; girls from other villages did not want to marry into and live in that village. The intelligence of the student population was known to be low (chil-

dren aged 10 had a mental development equivalent to those aged 7).

Iodized salt was introduced in 1978, and by 1982 the goiter rate had dropped to 4 percent. No cretins had been born since 1978. The attitude of the people changed greatly, with their approach to life being much more positive than before iodization. Average income increased from 43 yuan per capita in 1981 to 223 yuan in 1982 and 414 yuan by 1984 – the latter higher than the average per capita income of the district.

Before iodization, no family owned a radio in the village. But after iodization 55 families acquired television sets. Some 44 girls have arrived from other villages to marry boys in Jixian, and 7 men have joined the People's Liberation Army, whereas before men from the village had been rejected because of goiter. Such improvements in the level of living were mainly due to the correction of hypothyroidism by iodized salt.

The Demography of IDD

The prevention and control of IDD must be regarded as an urgent problem for the 1 billion of the world's peoples living in an iodine-deficient environment (Table IV.B.2.2). Of these, 350 million live in China and 200 million in India (*World Health Organization Report* 1990) (Map IV.B.2.1).

These populations are at risk of developing one or more of the iodine-deficiency conditions that can retard physical, social, and economic growth and development. Indeed, more than 5 million of the world's people are estimated to be suffering from mental retardation as gross cretins. But in addition,

Table IV.B.2.2. *Estimated prevalence of iodine-deficiency disorders in developing countries, by region and numbers of persons at risk (in millions)*

	At risk	Goiter	Overt cretinism
Africa	227	39	0.5
Latin America	60	30	0.3
Southeast Asia	280	100	4.0
Asia (other countries including China)	400	30	0.9
Eastern Mediterranean	33	12	–
Total	1,000	211	5.7

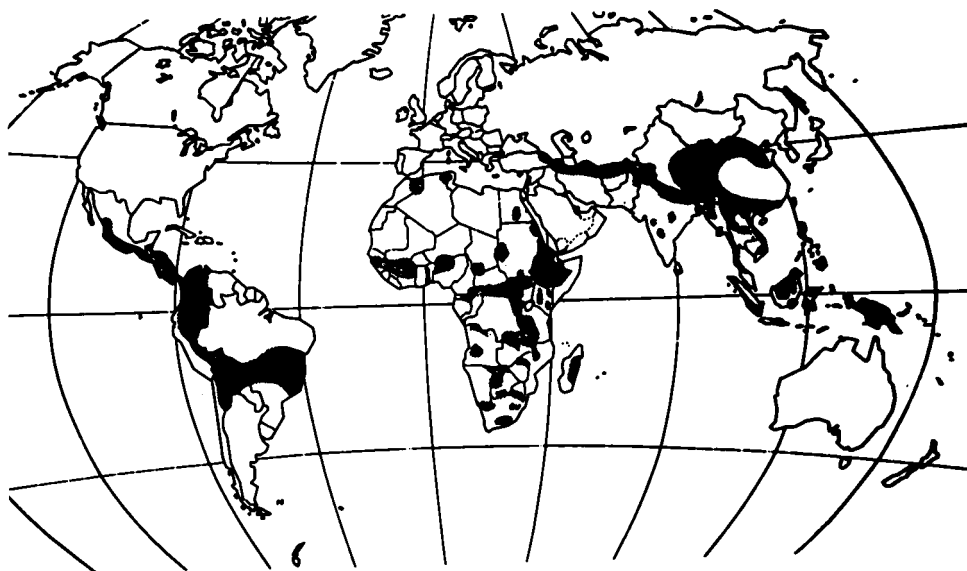
Source: From *World Health Organization Report* (1990).

three to five times this number suffer from lesser degrees of mental defect because their mothers were iodine-deficient during pregnancy (WHO 1990).

Iodine Technology for Prevention

Iodized Salt

Iodized salt has been the major preventive method used against iodine deficiency since the 1920s. However, there are serious difficulties in the production and quality maintenance of iodized salt for the millions that are iodine deficient. In Asia, the cost of iodized salt production and distribution at present is of the order of 3 to 5 cents per person per year. This



Map IV.B.2.1. The distribution of iodine-deficiency disorders in developing countries. Large populations in Asia provide two-thirds of the one billion at risk due to iodine deficiency of the soil. (From Hetzel 1989 and WHO 1990, with permission.)

must be considered cheap in relation to the benefits produced but there remain logistical problems of getting the salt to the iodine-deficient subjects. In addition, there may be problems with the iodine content – the salt may be left uncovered or exposed to heat, which leads to loss of iodine by volatilization.

Finally, there is the difficulty of getting people to consume the salt. Although the addition of iodine makes no difference to the taste of salt, the introduction of a new variety of the mineral to an area where it has been available, and its source familiar, is likely to be resisted. In the Chinese provinces of Xinjiang and Inner Mongolia, for example, the strong preference of the people for desert salt (of very low iodine content) led to a mass iodized oil injection program to prevent cretinism (Ma et al. 1982).

Iodized Oil by Injection

The value of iodized oil injection in the prevention of endemic goiter and endemic cretinism was first established in New Guinea with trials involving the use of saline injection as a control. Experience in South America and elsewhere has confirmed this value (Hetzel et al. 1980), and, in fact, the quantitative correction of severe iodine deficiency for a period of over four years by a single intramuscular injection (2 to 4 ml) has been demonstrated (Buttfield and Hetzel 1967). Iodized oil can be administered through local health

services, where they exist, or by special teams. In New Guinea, public health teams carried out the injection of a population in excess of 100,000, along with the injection of triple antigen. In Nepal 4 million injections have now been given. In Indonesia some 1 million injections were given between 1974 and 1978, together with the massive distribution of iodized salt (Djokomoeljanto, Tarwotjo, and Maspaitella 1983). A further 4.9 million injections have been given by specially trained paramedical personnel in the period 1979 to 1983. In Xinjiang in China, 707,000 injections were given by “barefoot doctors” between 1978 and 1981, and a further 300,000 to 400,000 injections were being given in 1982 (Ma et al. 1982). Iodized oil is singularly appropriate for the isolated village community so characteristic of mountainous endemic goiter areas. Iodized walnut oil and iodized soybean oil are more recently developed preparations and have been used on a wide scale in China since 1980. As with iodized salt, the striking regression of goiter following iodized oil injection ensures general acceptance of the measure (Figure IV.B.2.5).

In at-risk areas the oil should be administered to all females up to the age of 45 years, and to all males up to the age of 15 years and perhaps beyond. A repeat of the injection would be required in three to five years depending on dose and age. With children the need for a repeat injection is greater than with adults, and



Figure IV.B.2.5. Nodular goiter in a New Guinean (A) before and (B) three months after injection of iodized oil. The photos demonstrate the subsidence of goiter following the injection of iodized oil. (From Butterfield and Hetzel 1967, with permission.)

the dose should be repeated in three years if severe iodine deficiency persists (Stanbury et al. 1974).

Disadvantages of injections are the immediate discomfort produced and the infrequent development of abscesses at the site of injection. Sensitivity phenomena have not been reported (Hetzel et al. 1980), but a potential difficulty is disease transmission through syringes (hepatitis B or HIV infection). A major problem with injections is their cost, although the expense has been reduced with mass packaging to an order of magnitude similar to the costs of iodized salt, especially when the population to be injected is restricted to women of reproductive age and to children, and a primary health-care team is available (Hetzel 1983; SEARO/WHO 1985; Dunn et al. 1986).

Iodized Oil by Mouth

The effectiveness of a single oral administration of iodized oil for one to two years has been demonstrated in South America and in Burma. But recent studies in India and China reveal that oral iodized oil lasts only half as long as a similar dose given by injection (Dunn 1987b). However, oral administration avoids the hazard of transmitting hepatitis B or HIV infection through contaminated syringes. Moreover, oral administration of iodized oil to children could be carried out through baby health centers and schools. A 1 ml dose (480 mg) covers a period of 12 months at present.

Less Common Prevention Technologies

The use of other methods of iodization such as iodized bread and iodized water has not proceeded to the mass population level but may be indicated in special situations.

Iodized Bread

Iodized bread has been used in Holland and in Australia, and detailed observations are available from the Australian island of Tasmania. Since 1949 the Tasmanian population has received iodine dietary supplements in the form of weekly tablets of potassium iodide (10 mg) given to infants, children, and pregnant women through baby health clinics, schools, and antenatal clinics whenever possible. The prevalence of endemic goiter fell progressively during the next 16 years but was not eliminated. The failure to eliminate the disease completely was traced to a lack of cooperation by a number of schools in the distribution of the iodide tablets. The distribution through the child health centers to infants and preschool children was also ineffective because of the children's irregular attendance.

For this reason a decision was made to change the method of prophylaxis from iodide tablets to iodized bread. The use of potassium iodate up to 20 parts per million (ppm) as a bread improver was authorized by the National Health and Medical Research Council of Australia in May 1963, and the necessary legislation was passed by the Tasmanian Parliament in October 1964.

The effects of bread iodization were measured by a series of surveys of palpable goiter rates in schoolchildren. A definite decrease in the visible goiter rate was apparent by 1969. Yet studies of urinary iodide excretion and plasma inorganic iodide in May 1967 revealed no excessive intake of iodide. Correction of iodine deficiency was confirmed by evidence of a fall of 24-hour radioiodine uptake levels in hospital subjects as well as normal plasma inorganic iodine concentration and urine iodine excretion (Stewart et al. 1971).

Thus bread iodization was effective in correcting iodine deficiency in Tasmania in the 1960s. However, today there is a greater diversity of sources for dietary iodine. It is readily available from milk due to the use of iodophors in the dairy industry, and it is also available from ice cream due to the use of alginate as a thickener.

Iodized Water

Reduction in goiter rate from 61 percent to 30 percent with 79 percent of goiters showing visible reduction has been demonstrated following water iodization in Sarawak (Maberly, Eastman, and Corcoran 1981). Similar results have been obtained with preliminary studies in Thailand by Romsai Suwanik and his group at the Siriraj Hospital, Bangkok, and in Sicily (Hetzel 1989). Iodized water may be a more convenient preventive measure than iodized salt and there may be less likelihood of iodine-induced thyrotoxicosis. Certainly this method is appropriate at the village level if a specific source of drinking water can be identified.

The Hazards of Iodization

A mild increase in incidence of thyrotoxicosis has been described following iodized salt programs in Europe and South America and following the introduction of iodized bread in Holland and Tasmania. A few cases have been noted following iodized oil administration in South America. That no cases have yet been described in New Guinea, India, or Zaire is probably due to the scattered nature of the population in small villages and limited opportunities for observation. In Tasmania it was apparent that a rise in incidence of thyrotoxicosis accompanied a rise in iodine intake from below normal to normal levels due to iodized bread, which was finally introduced in April 1966 (Stewart et al. 1971). The condition can be readily controlled with antithyroid drugs or radio-iodine. Nonetheless, in general, iodization should be minimized for those over the age of 40 because of the risk of thyrotoxicosis (Stanbury et al. 1974; Hetzel 1983).

International Action Toward Elimination

Since 1978, there has been increasing concern about a lag in the application of our new knowledge of the iodine-deficiency disorders and their prevention

(Stanbury and Hetzel 1980; Hetzel 1983). A symposium at the Fourth Asian Congress of Nutrition in Bangkok in 1983 brought the problem to the attention of the international nutrition community with the preparation of a comprehensive report on the "Prevention and Control of Iodine Deficiency Disorders for the United Nations System" (*Lancet* 1983).

This report, which was subsequently published (Hetzel 1988b), indicated the need for the establishment of an expert group of scientists and other health professionals committed to bridging the wide gap between research and action. At a WHO/UNICEF Intercountry Meeting in March 1985, in Delhi, a group of 12 such experts agreed to found the International Council for Control of Iodine Deficiency Disorders (ICCIDD). In 1986 an International Board was established and the ICCIDD was formally inaugurated at Kathmandu, Nepal (*Lancet* 1986a).

The ICCIDD has welcomed as members all who have concern for and expertise in IDD and IDD control (*Lancet* 1986a). The main role of the ICCIDD is to make such expertise available to international agencies and national governments who have responsibility for the control of IDD, and so bridge the great gap between available knowledge and its application. At the inaugural meeting in Kathmandu, Nepal, a global review of the IDD problem took place along with an appraisal of current expertise in iodine technology and IDD control programs. This meeting led to the first ICCIDD monograph (*Lancet* 1986b; Hetzel, Dunn, and Stanbury 1987).

The ICCIDD is now a global, multidisciplinary group, consisting of some 300 members from 70 countries, and has a quarterly newsletter and a secretariat in Adelaide, Australia. Support since 1985 has been received from UNICEF (New York) and the Australian government and since 1991 from the Canadian government.

Since 1987, the ICCIDD has become fully active in facilitating interagency cooperation in the prevention and control of IDD by holding a series of interagency meetings and establishing regional interagency working groups (particularly involving WHO and UNICEF, but also some of the bilaterals). A Global Action Plan has now been endorsed by the UN agencies to take account of these developments. The purpose of the Global Action Plan is to provide global and regional support for the establishment and monitoring of effective national IDD control programs. It includes activities at the national, regional, and global levels.

At the national level, such activities include initial assessments, national seminars, communication packages, intersectoral planning with a National IDD Control Commission, evaluation, and monitoring with laboratory services. In Indonesia and China, International Working Groups have now been established by ICCIDD in collaboration with WHO and UNICEF.

At the regional level, the development of a series of regional IDD working groups provides for the necessary close working relationship between ICCIDD, WHO, and UNICEF. The IDD Task Force for Africa has been particularly successful in developing a coordinated strategy involving both multilateral and bilateral agencies.

At the global level, major functions include advocacy and the provision of public information, as well as the maintenance of a global monitoring system.

In view of the progress already achieved and the promising potential of current and planned national prevention and control programs, the goal of eliminating IDD as a major public health problem by the year 2000 was reaffirmed by the 1990 World Health Assembly (*World Health Organization Report* 1990). An escalation of funding now became an essential step in the achievement of this goal.

In September 1990, the World Summit for Children held at the United Nations in New York was attended by 71 heads of state and 80 other government representatives. The World Summit signed a Declaration and approved a Plan of Action that included the elimination of IDD as a public health problem by the year 2000.

This summit was followed in October 1991 by a conference entitled "Ending Hidden Hunger," which was a policy and promotional meeting on micronutrients including iodine, vitamin A, and iron. It was attended by multidisciplinary delegations from 55 countries with major IDD problems. These delegates were nominated by heads of state in response to an invitation by the director general of the World Health Organization, H. Nakajima, and the executive director of UNICEF, James Grant. At this meeting there was a firm commitment to the goal of eliminating IDD and vitamin A deficiency and reducing iron deficiency by one-third of 1990 levels. These various developments encourage the hope that very significant progress would be made toward the elimination of IDD during the last decade of the twentieth century with great benefits to the quality of life of the many millions affected (Hetzel 1989).

Basil S. Hetzel

Bibliography

- Bautista, S., P. A. Barker, J. T. Dunn, et al. 1982. The effects of oral iodized oil on intelligence, thyroid status, and somatic growth in school-aged children from an area of endemic goitre. *American Journal of Clinical Nutrition* 35: 127-34.
- Bleichrodt, N., I. Garcia, C. Rubio, et al. 1987. Developmental disorders associated with severe iodine deficiency. In *The prevention and control of iodine deficiency disorders*, ed. B. S. Hetzel, J. T. Dunn, and J. B. Stanbury, 65-84. Amsterdam.
- Boyages, S. C., J. K. Collins, G. F. Maberly, et al. 1989. Iodine deficiency impairs intellectual and neuromotor devel-

- opment in apparently-normal persons. *Medical Journal of Australia* 150: 676-82.
- Boyages, S. C., J.-P. Halpern, G. F. Maberly, et al. 1988. A comparative study of neurological and myxedematous endemic cretinism in western China. *Journal of Clinical Endocrinology and Metabolism* 67: 1262-71.
- Brush, B. E., and J. K. Altland. 1952. Goitre prevention with iodised salt: Results of a 30 year study. *Journal of Clinical Endocrinology* 12: 1380-8.
- Burgi, H., Z. Supersaxo, and B. Selz. 1990. Iodine deficiency diseases in Switzerland one hundred years after Theodor Kocher's survey: A historical review with some new goitre prevalence data. *Acta Endocrinologica* 123: 577-90.
- Burrow, G. N., ed. 1980. *Neonatal thyroid screening*. New York.
- Buttfield, I. H., M. L. Black, M. J. Hoffmann, et al. 1965. Correction of iodine deficiency in New Guinea natives by iodised oil injection. *Lancet* 2: 767-9.
- Buttfield, I. H., and B. S. Hetzel. 1967. Endemic goitre in eastern New Guinea with special reference to the use of iodised oil in prophylaxis and treatment. *Bulletin of the World Health Organization* 36: 243-62.
- Connolly, K. J., P. O. D. Pharoah, and B. S. Hetzel. 1979. Fetal iodine deficiency and motor performance during childhood. *Lancet* 2: 1149-51.
- Delange, F., F. B. Iteke, and A. M. Ermans, eds. 1982. *Nutritional factors involved in the goitrogenic action of cassava*. Ottawa.
- De Quervain, F., and C. Wegelin. 1936. *Der endemische Kretinismus*. Berlin.
- Djokomoeljanto, R., I. Tarwotjo, and F. Maspaitella. 1983. Goitre control program in Indonesia. In *Current problems in thyroid research*, ed. N. Ui, K. Torizuka, S. Nagataki, and K. Miyai, 403-5. Amsterdam.
- Dunn, J. T. 1987a. Alternatives to salt and oil for iodine supplementation. In *The prevention and control of iodine deficiency disorders*, ed. B. S. Hetzel, J. T. Dunn, and J. B. Stanbury, 135-40. Amsterdam.
- 1987b. Iodised oil in the treatment and prophylaxis of IDD. In *The prevention and control of iodine deficiency disorders*, ed. B. S. Hetzel, J. T. Dunn, and J. B. Stanbury, 127-34. Amsterdam.
- Dunn, J. T., E. A. Pretell, C. H. Daza, and F. E. Viteri, eds. 1986. *Towards the eradication of endemic goitre, cretinism, and iodine deficiency*. Washington, D.C.
- Ermans, A. M., N. M. Moulameko, F. Delange, and R. Alhuwalia, eds. 1980. *Role of cassava in the aetiology of endemic goitre and cretinism*. Ottawa.
- Fierro-Benitez, R., R. Cazar, J. B. Stanbury, et al. 1986. Long-term effect of correction of iodine deficiency on psychomotor and intellectual development. In *Towards the eradication of endemic goitre, cretinism, and iodine deficiency*, ed. J. T. Dunn, E. A. Pretell, C. H. Daza, and F. E. Viteri, 182-200. Washington, D.C.
- Follis, R. H., Jr. 1963. A pattern of urinary iodine excretion in goitrous and non-goitrous areas. *American Journal of Clinical Nutrition* 14: 253-68.
- Hetzel, B. S. 1983. Iodine deficiency disorders (IDD) and their eradication. *Lancet* 2: 1126-9.
- 1988a. Iodine deficiency disorders. *Lancet* 2: 1386.
- 1988b. *The prevention and control of iodine deficiency disorders*. Nutrition Policy Discussion Paper No. 3, ACC/SCN, Rome.
1989. *The story of iodine deficiency: An international challenge in nutrition*. Oxford and Delhi.
- Hetzel, B. S., J. Chevadev, and B. J. Potter. 1988. The brain in iodine deficiency. *Neuropathology and Applied Neurobiology* 14: 93-104.
- Hetzel, B. S., J. T. Dunn, and J. B. Stanbury, eds. 1987. *The prevention and control of iodine deficiency disorders*. Amsterdam.
- Hetzel, B. S., and G. F. Maberly. 1986. Iodine. In *Trace elements in human and animal nutrition*. Fifth edition, ed. W. Mertz, 139-208. New York.
- Hetzel, B. S., B. J. Potter, and E. M. Dulberg. 1990. The iodine deficiency disorders: Nature, pathogenesis and epidemiology. *World Review of Nutrition Dietetics* 62: 59-119.
- Hetzel, B. S., C. H. Thilly, R. Fierro-Benitez, et al. 1980. Iodised oil in the prevention of endemic goitre and cretinism. In *Endemic goitre and endemic cretinism*, ed. J. B. Stanbury and B. S. Hetzel, 513-32. New York.
- Hunziger, H. 1915. *Der Kropf, eine Anpassung an jodarme Nahrung*. Bern.
- Karmarker, M. G., M. G. Deo, N. Kochupillai, and V. Ramalingaswami. 1974. Pathophysiology of Himalayan endemic goitre. *American Journal of Clinical Nutrition* 27: 96-103.
- Kelly, F. C., and W. W. Snedden. 1960. Prevalence and geographical distribution of endemic goitre. In *Endemic goitre*, ed. World Health Organization, 27-333. Geneva.
- Kimball, O. P. 1937. Prevention of goitre in Michigan and Ohio. *Journal of the American Medical Association* 108: 860-3.
- Kochupillai, N., and C. S. Pandav. 1987. Neonatal chemical hypothyroidism in iodine deficient environments. In *The prevention and control of iodine deficiency disorders*, ed. B. S. Hetzel, J. T. Dunn, and J. B. Stanbury, 85-93. Amsterdam.
- Konig, P. 1981. "Myxedematous" and "nervous cretinism": A historical review. In *Fetal brain disorders*, ed. B. S. Hetzel and R. M. Smith, 229-42. Amsterdam.
- Lancet* (Editorial). 1983. From endemic goitre to iodine deficiency disorders. *Lancet* 2: 1121-2.
- 1986a. Inaugural meeting of the International Council for Control of Iodine Deficiency Disorders. *Lancet* 1: 1164.
- 1986b. Prevention and control of iodine deficiency disorders. *Lancet* 2: 433-4.
- Langer, P. 1960. History of goitre. In *Endemic goitre*, ed. World Health Organization, 9-25. Geneva.
- Li, J.-Q., and X. Wang. 1987. Jixian: A success story in IDD control. *IDD Newsletter* 3: 4-5.
- Ma, T., T. Lu, U. Tan, et al. 1982. The present status of endemic goitre and endemic cretinism in China. *Food and Nutrition Bulletin* 4: 13-19.
- Maberly, G. F., C. J. Eastman, and J. M. Corcoran. 1981. Effect of iodination of a village water-supply on goitre size and thyroid function. *Lancet* 2: 1270-2.
- Marine, D., and O. P. Kimball. 1921. The prevention of simple goitre in man. *Journal of the American Medical Association* 77: 1068-72.
- McCarrison, R. 1908. Observations on endemic cretinism in the Chitral and Gilgit Valleys. *Lancet* 2: 1275-80.
- McCullagh, S. F. 1963. The Huon Peninsula endemic. The effectiveness of an intramuscular depot of iodised oil in the control of endemic goitre. *Medical Journal of Australia* 1: 769-77.
- McMichael, A. J., J. D. Potter, and B. S. Hetzel. 1980. Iodine deficiency, thyroid function, and reproductive failure. In *Endemic goitre and endemic cretinism*, ed. J. B. Stanbury and B. S. Hetzel, 445-60. New York.
- Merke, F. 1984. *The history and iconography of endemic goitre and cretinism*. Lancaster. Translated from the German original, 1971, Bern.
- Pharoah, P. O. D., I. H. Buttfield, and B. S. Hetzel. 1971. Neurological damage to the fetus resulting from severe iodine deficiency during pregnancy. *Lancet* 1: 308-10.
- Pharoah, P. O. D., and K. J. Connolly. 1987. A controlled trial

- of iodinated oil for the prevention of cretinism: A long-term follow-up. *International Journal of Epidemiology* 16: 68–73.
- Pharoah, P. O. D., K. J. Connolly, R. P. Ekins, and A. G. Harding. 1984. Maternal thyroid hormone levels in pregnancy and the subsequent cognitive and motor performance of the children. *Clinical Endocrinology* 12: 265–70.
- Pharoah, P. O. D., K. J. Connolly, B. S. Hetzel, et al. 1981. Maternal thyroid function and motor competence in the child. *Developmental Medicine and Child Neurology* 23: 76–82.
- Pharoah, P. O. D., F. Delange, R. Fierro-Benitez, and J. B. Stanbury. 1980. Endemic cretinism. In *Endemic goitre and endemic cretinism*, ed. J. B. Stanbury and B. S. Hetzel, 395–421. New York.
- Pretell, E. A., T. Torres, V. Zenteno, and M. Comejo. 1972. Prophylaxis of endemic goitre with iodised oil in rural Peru. In *Human development and the thyroid gland - relation to endemic cretinism*, ed. J. B. Stanbury and R. Kroc. SEARO/WHO. *Iodine deficiency disorders in South East Asia*. 1985. SEARO Regional Papers, No. 10. Delhi.
- Stanbury, J. B., G. L. Brownell, D. S. Riggs, et al. 1954. *Endemic goiter: The adaptation of man to iodine deficiency*. Cambridge, Mass.
- Stanbury, J. B., A. M. Ermans, B. S. Hetzel, et al. 1974. Endemic goitre and cretinism: Public health significance and prevention. *WHO Chronicle* 28: 220–8.
- Stanbury, J. B., and B. S. Hetzel, eds. 1980. *Endemic goitre and endemic cretinism: Iodine nutrition in health and disease*. New York.
- Stewart, J. C., G. I. Vidor, I. H. Butfield, and B. S. Hetzel. 1971. Epidemic thyrotoxicosis in northern Tasmania: Studies of clinical features and iodine nutrition. *Australia and New Zealand Journal of Medicine* 3: 203–11.
- Thilly, C. H. 1981. Goitre et cretinisme endémiques: Role étiologique de la consommation de manioc et stratégie d'éradication. *Bulletin of the Belgian Academy of Medicine* 136: 389–412.
- Thilly, C. H., F. Delange, and J. B. Stanbury. 1980. Epidemiological surveys in endemic goitre and cretinism. In *Endemic goitre and endemic cretinism*, 157–84. New York.
- UNICEF Regional Office Report for South-Central Asia. 1984. Report on iodised salt. Delhi.
- Vidor, G. I., J. C. Stewart, J. R. Wall, et al. 1973. Pathogenesis of iodine induced thyrotoxins: Studies in Northern Tasmania. *Journal of Clinical Endocrinology and Metabolism* 37: 901–9.
- WHO (World Health Organization). 1960. *Endemic goitre*. Monograph Series No. 44. Geneva.
- World Health Organization Report to World Health Assembly. 1990. Geneva.

IV.B.3 Iron

Iron has played a critical role in the evolution of life. The ancient Greeks, believing iron to be a special gift sent to earth by one of the gods, named it *sideros*, or star (Liebel, Greenfield, and Pollitt 1979). As the second most common metal, iron accounts for 5 percent of the earth's crust; it is also found in both sea- and freshwater (Bernat 1983). Scientists believe that the earth's atmosphere was originally a reducing one with

very low oxygen pressure. As a result, large amounts of reduced iron would have been available for living organisms (Bothwell et al. 1979). Iron is an essential element for all organisms, with the possible exception of some *Lactobacillus* (Griffiths 1987; Payne 1988). In animals, the processes of DNA replication, RNA synthesis, and oxygen and electron transport require iron. Today most iron in the environment exists in an oxidized state and is less available to organisms. However, the problems of extracting insoluble iron have been overcome during evolution. A variety of sophisticated mechanisms have evolved that are specific to different kingdoms and/or different species (e.g., mechanisms plants use to be able to live in acidic or iron-poor environments) (Bothwell et al. 1979). Such mechanisms in animals include iron complexing agents, which transport iron and deliver it to cells, and low-molecular-weight compounds, such as fructose and amino acids, that reduce iron into a soluble form (Griffiths 1987; Simmons 1989: 14).

Metabolic processes within humans involve the presence of free radicals, that is, substances that are reactive because of instability in the arrangement of electrons (Wadsworth 1991). Iron may be present as a free radical. Such instability makes iron highly likely to donate or accept electrons. As a result, iron is versatile and able to serve a number of functions within cells. These functions include acting as a catalyst in electron transport processes and serving as a transporter of oxygen. Iron is a key component of hemoglobin, the oxygen carrier found in red blood cells. It is involved in many other extracellular processes as well (Woods, DeMarco, and Friedland 1990). Iron also is required for collagen synthesis, the production of antibodies, removal of fats from the blood, conversion of carotene to vitamin A, detoxification of drugs in the liver, and the conversion of fuel nutrients to energy (Long and Shannon 1983). In addition to its importance in the maintenance of normal metabolic processes, iron involvement in pathological change and initiation of disease is a critical facet of host defense.

Iron occurs in the human body in two states: circulating in blood and in storage. The absolute amount of iron in the body is actually quite small, despite its critical importance in metabolism. A healthy adult has an average of between 4 and 5 grams (g) of iron with a range of 3 to 6 g (Bernat 1983). Approximately 30 percent of all iron in the body is in storage. Most iron utilized in the body is needed by bone marrow to make new red blood cells. After a life span of about 120 days, red blood cells are destroyed and the released hemoglobin is broken down. Liberated iron from the hemoglobin is stored in macrophages in the liver, spleen, and kidney for future use.

Most iron in a healthy person (2.5 to 3 g) is found in red blood cells in the form of hemoglobin (Moore 1973; McLaren, Muir, and Kellermeyer 1983). Small amounts of iron, approximately 0.13 g, occur in myo-

globin, a protein in muscle tissues. Only 0.008 g of iron is located in heme tissue enzymes, 0.5 g is in nonheme enzymes, and about 0.004 g is bound to the iron-binding protein transferrin found in blood (Arthur and Isbister 1987).

Between 0.7 and 1.5 g of iron is found in storage (Bezkorovainy 1989). It occurs in either a diffuse soluble form as ferritin, an intracellular protein, or as hemosiderin, an insoluble aggregate form of ferritin (Bezkorovainy 1989). Ferritin is formed when apoferritin (an iron-free protein) combines with freed iron from the breakdown of hemoglobin in senescent red blood cells (Simmons 1989: 14). Basically, ferritin is an apoprotein shell enclosing a core of iron that consists of 24 subunits of either the H or the L type (Worwood 1989). The two types of subunits are of slightly different size and electrical charge. Different organs have various percentages of one or the other type of subunits. For instance, H subunits are predominantly found in ferritin located in the heart and L subunits are predominantly found in ferritin located in the liver. It has been suggested that those organs that have a high iron content, such as the liver and spleen, have a predominance of the L subunits, whereas those with less iron, such as heart, intestine, pancreas, and placenta, have more of the H subunits (Bezkorovainy 1989: 53). Therefore, the L subunit dominated ferritin is thought to be concerned primarily with long-term iron storage whereas the H-dominated ferritin is concerned primarily with preventing iron overload (Bezkorovainy 1989: 53). Although the majority of ferritin is located in the organs that make up the reticuloendothelial system, a small amount exists in plasma. Most ferritin molecules contain only 2,000 atoms of iron (Fairbanks and Beutler 1988). However, up to 4,500 atoms of iron can be incorporated into the ferritin internal cavity. Iron atoms enter the ferritin protein coat through tiny channels or pores, which are formed by the positioning of subunits (Fairbanks and Beutler 1988; Worwood 1980: 204; Worwood 1989). Hemosiderin represents the end point of the intracellular storage iron pathway. Hemosiderin granules are actually denatured, partly degraded, aggregated molecules of ferritin (Bezkorovainy 1989).

Balance is the crucial feature of human iron metabolism. Too much iron is toxic and too little can produce severe anemia (Arthur and Isbister 1987). In fact, much of the toxicity of iron appears to be related to its propensity to form unstable intermediates with unpaired electrons or free radicals (Griffiths 1987). Lethal accumulations of iron can occur in most organs of the body but the liver, spleen, pituitary, and heart are particularly vulnerable to excess iron (Weinberg 1984; Stevens et al. 1988; Cook 1990; Kent, Weinberg, and Stuart-Macadam 1990, 1994). The immune system can also be compromised by excess iron. For this reason it is extremely important that iron be chelated (bound) to prevent uncontrolled free-radical

reactions. In normal animals and humans, iron is almost always bound to proteins, leaving only an extremely low concentration of free iron; in fact, in most biological systems iron is bound to complexing agents. Over 99 percent of iron in plasma is chelated to transferrin to prevent uncontrolled free-radical reactions (Griffiths 1987). However, in individuals with hyperferremia (iron overload), as much as 35 percent of iron may not be transferrin-bound, causing iron to accumulate in the liver, spleen, and heart (Weinberg 1989: 10).

By contrast, insufficient iron can produce severe anemia that impairs the quality of life and may eventually lead to cardiac and respiratory failure (Weinberg 1989). The body maintains an equilibrium in part through the transfer of circulating iron to storage, as well as through the conversion of free iron to bound iron. The key to body iron supplies is recycling; only a very small amount is absorbed from food or lost through excretion (sweat, feces, urine, or menstruation). Iron is conserved in a nearly closed system. Each iron atom cycles repeatedly from plasma and extracellular fluid to the bone marrow, where it is incorporated into hemoglobin, and then to the blood, where it circulates for approximately 120 days within erythrocytes (red blood cells). Afterward, iron is released to plasma. This is accomplished by phagocytes (cells that engulf and then consume debris and foreign matter) of the liver or spleen (the reticuloendothelial system¹) that digest hemoglobin and destroy old erythrocytes. The process releases iron to plasma, which continues the cycle (Fairbanks and Beutler 1988: 195). Within each cycle, small amounts of iron are transferred to storage sites, and small amounts of storage iron are released to plasma. In addition, small amounts are absorbed by the intestinal tract from ingested food containing iron and small amounts are lost through sweat, urine, feces, or blood (Fairbanks and Beutler 1988: 195). Consequently, little dietary iron is needed to replace iron loss.

Iron Requirements in Humans

Because most needed iron is recycled, humans actually require very small quantities to maintain health. For example, healthy 70-kilogram (kg) adult males lose approximately 1 milligram (mg) of iron daily through excretion by the skin and the gastrointestinal and urinary tracts (Bothwell et al. 1989).² Required replacements are 1 mg daily. Children lose only about 0.5 mg daily but their required replacement is approximately 1 mg daily to compensate for the needs of growth. Early estimates of iron loss among women of reproductive age were high, an average of 2 to 3 mg per day, resulting in an equal replacement requirement (Hoffbrand and Lewis 1981; Arthur and Isbister 1987). However, more recent and systematic studies call such estimates into question. A Swedish study showed that the average woman loses between

0.6 and 0.7 mg through menstruation; 95 percent of women lose an average of *less* than 1.4 mg per day (Fairbanks and Beutler 1988: 205). Scientists are beginning to recommend that average menstruating women absorb only 1.4 mg iron daily to replace losses (Monsen 1988: 786).

Absorption of iron occurs mainly in the duodenum and jejunum, although all parts of the small intestine and even the colon may be involved. Little is known about the precise mechanisms by which the body's need for iron is communicated to the mucosal cell of the intestine (Cook 1990). But however it is achieved, the body is able to adapt to a wide range of iron requirements and intakes by modifying the rate of gastrointestinal absorption. Absorption of iron varies according to health, physiological status, iron stores, age, and sex. Studies have shown that the absorption of iron from an adequate diet might range from a fraction of a milligram to 3 or 4 mg a day, depending on body iron content. For example, hyperferremia, or iron overload, is associated with decreased iron absorption from dietary sources. Blood loss and true iron deficiency result in increased iron absorption. Absorption also varies according to the iron source. Iron-deficient males absorb 20 percent more non-heme iron and 21 percent more heme iron than iron-replete males (Cook 1990: 304). The same adaptability applies to iron loss. Normal males lose about 0.9 mg of iron per day. In contrast, hypoferremic males lose only about 0.5 mg per day and hyperferremic males lose about 2.0 mg per day (Finch 1989).

According to James Cook (1990), the three major determinants of body iron in healthy individuals are physiological iron demands, dietary supply of available iron, and adaptation. It is the dietary supply of available iron that is most often focused upon by the medical profession and nutritionists. Dietary iron supply is actually determined by three main factors: total iron intake, content of heme iron, and bioavailability of nonheme iron (Cook 1990). However, actual absorption of iron is a function of the adaptability of iron metabolism and an individual's iron status. Dietary iron comes in two forms - heme and non-heme. The largest percentage of dietary iron is inorganic, or nonheme, which is absorbed primarily in the duodenum. Its absorption is influenced by the presence of inhibitors and enhancers found in ingested foods. The most important inhibitors include tannin (found in tea), coffee, phytates in bran, calcium phosphate, egg yolk, polyphenols, and certain forms of dietary fiber. Although bran and phytate in the diet do inhibit iron absorption, there is speculation that high bran and phytate intake may induce changes in the intestines or its microflora over a prolonged period and reduce the inhibitory effect of these substances (Cook 1990). Enhancers include meat and organic acids such as citric acid, lactic acid, alcohol, certain spices and condiments in curry powder, and ascorbic acid (Bothwell et al. 1989).

Unlike nonheme iron, heme iron is absorbed directly into the intestinal cell as an intact porphyrin complex and is unaffected by other components of the diet. Heme iron usually accounts for only 10 to 15 percent of the total daily intake of iron; however, the absorption of heme iron is relatively high and it can account for as much as 25 percent of the iron absorbed from the daily diet (Hallberg 1980: 118).

Most Westerners, who ingest some kind of meat almost daily, easily obtain sufficient amounts of iron to replace daily losses. In fact, in some cases, dietary consumption is four to seven times greater than the amount of iron needed. They also consume iron through fortified foods, such as most wheat products or wine and cider, which can add 2 to 16 mg or more per liter (Fairbanks and Beutler 1988). A study of military personnel showed that 20 to 30 mg of iron per day was ingested, in part because food was cooked in iron containers (Fairbanks and Beutler 1988: 194-5).

Several factors make it difficult to quantify the bioavailability of iron in the body. Uncertainties concerning the nature of the interactions that occur between enhancers and inhibitors make it difficult to determine the true bioavailability of iron from various foods or diets. Compounding the problem are misconceptions that result from not considering a food source within its natural context. Often quoted is the belief that milk is a poor source of iron. Although this is true for humans drinking bovine milk, it is not true for human infants drinking human breast milk. There is a very high bioavailability of iron in human breast milk for human infants. As much as 40 percent of iron in human milk is absorbed by human infants, whereas less than half that amount is absorbed from bovine milk (Pochedly and May 1987). Iron availability of milk is species-specific, and calves absorb significantly more from bovine milk than they would from human milk (Picciano and Guthrie 1976).

Not only is Western dietary intake of iron high; the Recommended Dietary Allowance (RDA) of the United States Food and Nutrition Board is high: 18 mg per day for premenopausal women and 10 mg per day for men and older women (Eaton, Shostak, and Konner 1988). This high RDA leads to perceived high levels of iron deficiency in populations throughout the world. Medical and government policy has traditionally encouraged more iron consumption over less. Accordingly, iron supplements are routinely administered to pregnant women and infants. Iron supplementation on a massive scale has been carried out in the United States since the 1940s when the Food and Nutrition Board of the National Academy of Science endorsed fortification of wheat flour with iron as well as vitamins. It is now standard practice to fortify a broad range of foods, including many of the items available on the shelves of supermarkets around the world. In fact, the only infant formula available to disadvantaged women participating in the American WIC (Women, Infants, and Children)

program is one fortified by iron; non-iron-fortified formula is prohibited (Kent et al. 1990). Yet iron is naturally found in almost all foods.

Particularly rich sources of iron include organ meats, red meat, blackstrap molasses, cocoa, oysters, clams, and dried beans (Long and Shannon 1983). Water also can contain iron; that from wells and boreholes can have more than 5 mg of iron per liter (Fairbanks and Beutler 1988: 195). This is in addition to the large amount of iron that can be absorbed by food cooked in iron containers. Even if iron were nonexistent in a person's diet, which is very unlikely, it would take years to produce true dietary iron deficiency without concomitant blood loss that depletes body iron stores: "[I]ron deficiency is almost never due to dietary deficiency in an adult in our community. A diagnosis of dietary iron deficiency in an adult male or postmenopausal female usually means that the site of blood loss has been missed" (Arthur and Isbister 1987: 173). In fact, occult gastrointestinal blood loss has been detected in nearly 50 percent of infants misclassified as dietary-induced iron deficient (Fairbanks and Beutler 1988: 195). In addition, many cases of the anemia of infection/inflammation caused by chronic diseases have been mistaken for iron-deficiency anemia.

Measurement of Iron in Humans

Several blood indices are used to measure iron but only the most common ones are briefly described here (Table IV.B.3.1). One of the most reliable methods is bone marrow aspiration and biopsy, which permits a direct measurement of the body's iron stores as indicated by the amount of ferritin and hemosiderin present. However, because of the discomfort to the subject, the possibility of introducing infection, and the costs involved, the procedure is usually conducted only when other, less invasive measurements indicate a need for a highly accurate assessment.

In past studies hemoglobin and hematocrit (a measurement of the packed cell volume) were employed to ascertain iron sufficiency. The widespread use of automated electronic blood cell counters, such as the Coulter counter, makes hemoglobin/hematocrit values easily obtainable and therefore frequently used

indices to assess iron adequacy. In addition, the relative ease of determining these indices make them popular measurements in situations where automated machines do not exist. Unfortunately, a number of variables interact with and influence these indices, making them poor measures of iron status. Hemoglobin/hematocrit measurements can be unreliable because they are affected by cultural factors, such as smoking, and environmental factors, such as altitude, and the iron content of drinking water and soil (Kent et al. 1990).

Serum iron is also not a very reliable measure of iron status. It, like hemoglobin and hematocrit, is influenced by a number of factors, including a woman's menstrual cycle, the presence of infection, and the time of day blood is drawn (there is a diurnal cycle wherein serum iron values are as much as 30 percent higher in the morning than in late evening [Fairbanks and Beutler 1988: 200]). Transferrin is another component of blood measured to investigate the iron status of a person. Transferrin is a plasma iron-binding protein. One molecule has two iron-binding sites, each of which can bind one atom of ferric iron at each receptor site (Noyes 1985). Although originally thought to be similar, the two iron-binding sites more recently have been recognized as not chemically or physically equivalent for reasons not well understood (Bezkorovainy 1989). The primary function of transferrin is to transport iron from the intestinal tract, spleen, and liver to sites such as the bone marrow for hemoglobin synthesis, to macrophages for storage, to the placenta for fetal needs, or to other cells for iron-containing enzymes (Fairbanks and Beutler 1988: 200).

Most transferrin, between 70 and 90 percent, is transported to bone marrow (Fairbanks and Beutler 1988: 200). Total iron-binding capacity (TIBC) is the capacity of transferrin to bind iron, and it represents the highest amount of iron that the plasma (i.e., serum) can bind (Bezkorovainy 1980: 10; Fielding 1980: 15). In humans, only about one-third of the transferrin iron-binding sites are occupied by iron; the rest are free (Bezkorovainy 1989b). Transferrin saturation, the amount of transferrin that is saturated with iron (calculated by dividing serum iron value by the

Table IV.B.3.1. *Normal hematological values for the more common iron indexes (norms vary slightly according to measurement instrument and source)*

	Hemoglobin (g/dl)	Hematocrit (%)	Transferrin saturation (%)	Serum ferritin (μg/l)	Serum iron (μmol/l)	Transferrin (g/dl)	Red blood cell (μl)	Vitamin B ₁₂ (pg/l)	Serum folate (ng/ml)	Red cell folate (ng/ml)
Adults										
Males	14–18.2	40–54	16–50	12–250	12.6–27	1.9–4.3	4.4–6.0	200–1,100	3–14	160–600
Females	12–15	38–47	16–50	12–200	12.6–27	1.9–4.3	4.2–5.5	200–1,100	3–14	160–600
Children ^a	11–14.8	33–44.8	16–50	7–140	10.0–27	1.9–4.3	3.4–5.6	200–1,100	3–14	160–600

^aValues for children are age-dependent.

total iron-binding capacity), is usually considered a better assessment of iron metabolism than is transferrin, serum iron, or total iron-binding capacity alone. Nevertheless, transferrin saturation values are affected by chronic disease and inflammation, in addition to insufficient diet, blood loss, and various types of iron overload. Therefore, by themselves, these measurements are not reliable in determining the etiology of an anemia or iron overload. Furthermore, "transferrin measurements are relatively insensitive, as the degree of change with variation in iron stores is small relative to assay variability. Serum ferritin measurement is a key indicator of iron stores" (Cook and Skikne 1989: 350). In fact, and contrary to conventional practices, physicians are now suggesting different methods to measure iron status: "Traditional tests of serum iron and TIBC may be useful but currently are not recommended for indirect measurement of iron stores [i.e., measurement in the absence of a bone marrow biopsy]" (Beissner and Trowbridge 1986: 88-90). More recently, it has been noted that "a wide range of values for transferrin saturation and total iron-binding capacity is consistent with either iron deficiency anemia or anemia of chronic disease. . . . The two disorders can be distinguished by the serum ferritin level, iron stain of a bone marrow specimen for hemosiderin, or both" (Farley and Foland 1990: 92).

Red blood cell counts are routinely calculated by most automated blood cell counters. Red blood cells are often malformed with reduced life spans in various types of anemia and are associated with ineffective erythropoiesis (or red cell production) (Jacobs and Worwood 1982: 175). However, red blood cells usually have a normal life span in most iron overload disorders without severe liver disease (McLaren et al. 1983). At one time it was thought that red blood cell size might be useful in distinguishing iron deficiency from the anemia of chronic disease. But recent studies have disproved this proposition (e.g., Thompson et al. 1988; Osborne et al. 1989).

Mean corpuscular volume (MCV) reflects the capacity of the cell. A decreased cell size can signify a hemoglobin synthesis disorder, but a smaller size is also associated with both iron-deficiency anemia and anemia of chronic disease/inflammation and other iron disorders, such as thalassemia and sideroblastic anemia. Moreover, elevated MCV can result from a variety of disorders other than infection, including idiopathic ones. Although widely used because it is measured as part of the automated complete blood

cell count, the main disadvantage of MCV is that it does "not distinguish true iron deficiency from secondary iron-deficient erythropoiesis due to inflammatory or neoplastic disease" (Cook and Skikne 1989: 351). As a result, a number of researchers contend that MCV can provide only a crude measure of iron status (Labbe and Finch 1980: 52).

Storage iron is held within the ferritin molecule. Experiments show that serum ferritin is the most noninvasive diagnostic indicator of iron status. The reliability of serum ferritin surpasses that of free erythrocyte protoporphyrin, which in the past was thought by some to be more diagnostic in discriminating the etiology of anemia (Zanella et al. 1989). Serum ferritin is a particularly sensitive measurement because it reflects changes in iron stores before they are completely exhausted or, in the case of overload, increased (e.g., Finch and Huebers 1982; Cook and Skikne 1989). There are now numerous studies that show that serum ferritin is a reliable measure of iron stores, second only to actual bone marrow aspirations (e.g., Thompson 1988; Burns et al. 1990; Guyatt et al. 1990).

These hematological measurements together present a reliable view of the iron status of an individual. However, when all are not available, serum ferritin, in conjunction with a measure of the amount of circulating iron, is minimally necessary for reliable interpretations.

Iron and Infection

One of the most exciting frontiers in iron studies concerns the body's efforts to diminish iron content in response to disease. Hypoferremia (or low circulating-iron level) is associated with an array of chronic diseases and inflammatory responses, including neoplasia (cancer), bacterial and parasitic infections, and rheumatoid arthritis (Kent 1992). The associated anemia of chronic disease is characterized by subnormal hemoglobin, subnormal serum iron, subnormal transferrin saturation levels, and normal to elevated serum ferritin levels (Table IV.B.3.2).

Hypoferremia in reaction to chronic diseases and inflammations has been noted in experimental animals as well as in humans. Studies show that within 24 hours after exposure to pathogenic microbes, serum iron levels fall and a resistance to otherwise lethal doses of a variety of pathogens is produced (Payne 1988). There is even a slight drop in

Table IV.B.3.2. *Comparison of laboratory values of anemia of dietary iron deficiency and anemia of chronic disease*

Anemia	Hemoglobin	Serum iron	Transferrin saturation	Serum ferritin	ESR
Iron deficiency	Decreased	Decreased	Decreased	Decreased	Normal
Chronic disease	Decreased	Decreased	Decreased	Normal-Raised	Raised

hemoglobin levels after smallpox revaccination, which is thought to represent a rather innocuous procedure (Reizenstein 1983).

The body reduces iron levels by preventing macrophages from releasing the metal from storage to transferrin and by reducing intestinal iron absorption. Transferrin saturation levels drop concomitantly as iron is sequestered in macrophage storage. Serum ferritin levels increase, serving as an indirect measure of body iron stores. The reduction of available iron affects microorganisms and rapidly multiplying neoplastic (cancer) cells because they require iron for proliferation. However, they either cannot store the element in a nontoxic form, as is the case for a number of pathogens, or they require larger amounts of iron for rapid proliferation than they possess, as is the case for neoplastic cells. Microorganisms requiring host iron include bacteria, fungi, protozoan parasites, and neoplastic cells.

Microinvaders acquire iron from their hosts in several ways. Perhaps most intriguing is acquisition through the production of small iron-binding compounds called siderophores. These iron chelators, or "iron scavengers," seek out iron in the host and donate it to the microorganism that produced them. In addition, some siderophores may serve as iron donors for another strain or species (Bezkorovainy 1980: 305–6). Siderophores compete with the host for the iron-binding proteins. The body responds by producing fever, which reduces the ability of some bacteria to produce siderophores (Lee 1983; Weinberg 1984). That is, fever serves as a countermeasure to microorganism proliferation by inhibiting their ability to extract iron from the host. The resulting hypoferrremia prevents pathogens from obtaining sufficient quantities of growth-essential iron; they therefore fail to multiply or multiply with considerable difficulty (Kochan 1973: 22). Studies have documented that "a single episode of fever and/or inflammation in man appears to restrict the release of iron from effete red blood cells by the reticuloendothelial system, leading to a decrease in serum iron concentration and to stimulation of the production of ferritin [to store iron] for a prolonged period" (Elin, Wolff, and Finch 1977: 152).

It has long been recognized that an important deterrent to pathogens in infants is provided by lactoferrin, an iron-binding protein that occurs in breast milk. Lactoferrin binds free iron and thereby deprives microorganisms of necessary iron; it also has a propensity to interact with or complement specific antibodies (Bullen, Rogers, and Griffiths 1974). When iron-binding proteins become saturated with iron there is an increase in the amount of free iron (i.e., nonbound iron that is more accessible to pathogens). Both in vitro and in vivo experiments have shown that the bacteriostatic properties of lactoferrin, for example, are abolished when iron-binding proteins are saturated with iron, making the body more vulner-

able to microorganism invasion (Pearson and Robinson 1976).

Understanding hypoferrremia as a defense against disease allows us to understand the geographical distribution of anemia. George Wadsworth (1975), for example, noted that levels of hemoglobin in women living in temperate climates usually exceed 10 g per 100 milliliters (ml), whereas usual levels in women living in tropical climates may be only 8 g per 100 ml. The etiology of hypoferrremia provides a clue for this otherwise puzzling observation. Parasitic and bacterial infections are endemic in tropical environments (e.g., Goodner 1933). The higher prevalence of infection has been related to the increased total complexity and diversity of the tropical ecological system (Dunn 1968). Hypoferrremia is produced by the body in an attempt to ward off the numerous pathogenic insults associated with tropical environments. However, in these cases, and particularly when exacerbated by blood loss and nutrient malabsorption that often accompany protozoan and bacterial infections, hypoferrremia might develop to such a severe state that it becomes a problem in and of itself (Stuart-Macadam 1988; Kent and Weinberg 1989; Kent et al. 1990: 67).

The role of hypoferrremia as a nonspecific body defense against disease is not completely accepted, although microbiologists in particular have long recognized the role of iron in host defense. As early as 1868 Armand Trousseau noted that iron supplementation reactivated quiescent pulmonary tuberculosis (Weinberg 1984; Keusch and Farthing 1986). Research in 1944 showed that iron-binding proteins in egg whites and serum were involved in the inhibition of *Shigella* dysentery (Schade and Caroline 1944). Starving humans who are susceptible to some diseases are immune to others in which microorganisms must exact iron from their hosts to obtain the amount needed to proliferate. For example, refeeding programs in Niger reactivated malaria infections (Murray et al. 1976). Similar refeeding programs in Somalia resulted in a recrudescence of a number of different infections, including parasitic and bacillary dysentery, acute infectious hepatitis, brucellosis, tuberculosis, and others (Murray et al. 1976; Murray and Murray 1977). Moreover, a study of two groups of Kenyan Turkana pastoralists indicated lower iron levels enhanced host immunity. One group drank primarily milk; the other group drank milk and consumed over 150 g of fish per day, ingesting almost a third more iron than the non-fish-eating group. Those who ate fish were not anemic but were plagued by malaria, brucellosis, molluscum contagiosum, common warts, diarrhea, and *Entamoeba histolytica*. The non-fish-eating Turkana were slightly anemic but had significantly lower incidences of these diseases (Murray, Murray, and Murray 1980).³

There is now a body of literature based on in vivo and in vitro experiments that demonstrates the role

of iron in disease (e.g., Weinberg 1966, 1974, 1977, 1984, 1990; Bullen et al. 1974; Masawe, Muindi, and Sway 1974; Strauss 1978; Oppenheimer et al. 1986; Bullen and Griffiths 1987; Crosa 1987; Griffiths and Bullen 1987; Kluger and Bullen 1987; Selby and Friedman 1988; Stevens et al. 1988; Kent et al. 1990, 1994). Even so, a large percentage of the public and even those in the health-care community are not aware that in certain circumstances diminished iron is not a pathological disease state but a physiological defense.⁴ They are unaware that the current focus on iron fortification of food and on iron supplementation preparation is misdirected and potentially counterproductive to the body's attempt to ward off illness.

Supplying iron-fortified foods or vitamins to developing countries and disadvantaged minorities in Western nations without ascertaining the prevalence of dietary-induced, versus disease-induced, anemia may be harming the very people the food is intended to help. The practice of indiscriminate food fortification is putting these populations at risk for contracting a variety of diseases, including cancer (Finch 1989). Studies show that the high iron content of most Western diets particularly affects men and postmenopausal women, who, as a result, suffer more from certain types of neoplasia and from myocardial infarction. In fact, recent treatment uses iron chelators to bind iron at the location of inflammation, such as at the joints in rheumatoid arthritis, where it is thought iron causes tissue damage, among other problems (Biernacki et al. 1988). In addition, drugs such as aspirin prevent heart attacks, at least in part, by causing chronic intestinal bleeding, which reduces iron levels (Sullivan 1989; also see Arthur and Isbister 1987 for discussion of aspirin and other anti-inflammatory drugs that can cause occult blood loss).

Iron and Heart Disease

Despite extensive epidemiological studies to identify risk factors for coronary disease, many unanswered questions remain. Individuals who have multiple risk factors often never develop cardiac events whereas others who would seem to be low risk experience cardiovascular problems (Sullivan 1989). Therefore, other possible factors, such as iron levels, should be examined for a contributory role in causing heart disease.

Some studies suggest a link between cardiac diseases and iron levels (Sullivan 1989; Salonen et al. 1992). Excess iron, as occurs in beta-thalassemia and chronic iron overload, has been associated with ischemic heart disease and heart attacks (Leon et al. 1979). Men, with more iron than premenopausal women who regularly lose iron through menstruation, also have more heart disease (Weinberg 1983; Sullivan 1989). Postmenopausal women, who have more iron than premenopausal women, have a higher risk of heart disease, although, of course, they constitute an older group. But women taking oral contra-

ceptives, which reduces the amount of menstrual bleeding, are also more prone to heart disease (Sullivan 1983; Frassinelli-Gunderson, Margen, and Brown 1985; Colditz et al. 1987).

Not all cardiologists agree that the association between iron and cardiac disease is causal. However, J. L. Sullivan (1989) and others (Salonen et al. 1992) have suggested that high serum ferritin values, which measure the amount of iron in storage, are a powerful predictor for heart disease. Common iron chelators used clinically to reduce heart attacks, such as deferoxamine and aspirin, cause increased bleeding times and occult blood loss. Moderate to heavy fish consumption causes occult bleeding, thereby reducing iron levels and the risk of heart disease (Kromhout, Bosschieter, and Coulander 1985; Herold and Kinsella 1986; Houwelingen et al. 1987). The consequent reduction in iron levels among consumers of fish and iron chelators or drugs that promote bleeding could be responsible for the reduction of heart disease. This is in contrast to persons who consume large amounts of red meat that contain heme iron, which is more readily absorbed than nonheme iron found in vegetables. Another interesting association between available iron and heart attacks is based on the circadian rhythm of the body. Serum iron levels have diurnal patterns with an increase in the early morning. Heart attacks also occur three times more frequently during this time of day (Sullivan 1989: 1184).

Iron and Parasites

Parasites are a serious threat to the health of rural communities in the tropics; over 90 percent of this population are infected with at least one and often several nematode species, including hookworm, roundworm, and whipworm (Behnke 1987). There are basically three ways in which parasites can cause anemia. One is the anemia of chronic disease discussed earlier. A second way is by competing for the same nutrients as the host, including iron. A third way is by causing actual blood loss. Hookworm (*Necator* sp.), in particular, can cause heavy blood loss and, as a result, anemia. Hookworm infestations have been found in 12 percent to 95 percent of study populations throughout the tropical regions of the world - from the Amazon rain forest to Southeast Asia (Florentino and Guirriec 1984: 65). Malaria parasites are also dependent on iron from their hosts. The distribution of the sickle trait and heterozygous forms of thalassemia, which cause anemia, provide some protection against the malaria parasite.

Iron and Cancer

Another exciting path iron studies are exploring concerns the examination of the role of iron in neoplastic disease. Neoplastic cells require more iron than normal cells because of their rapid proliferation. In an attempt

to accommodate their increased need for iron, these cells extract iron from their host. In reaction, the body attempts to prevent their proliferation by producing hypoferremia. Anemia is a commonly associated condition of neoplasia of various types (Miller et al. 1990). The type of anemia that usually occurs with neoplasia is that of chronic disease/inflammation as indicated by elevated serum ferritin levels. For example, patients with acute leukemia, Hodgkin's disease, multiple myeloma, and other malignancies have elevated to highly elevated serum ferritin concentrations that tend to become greater as the disease progresses (Bezko-rovany 1980; Worwood 1980).

Males with higher iron stores than premenopausal women have a significantly increased risk of cancer of the colon, bladder, esophagus, and lung. Women with higher levels of iron because they are postmenopausal or are hyperferremic also have a greater risk of cancer (Selby and Friedman 1988; Stevens et al. 1988). Very high serum ferritin levels are associated with other types of neoplasia, including malignant histiocytosis, primary hepatocellular carcinoma (liver cancer), and multiple malignancies (Hann et al. 1989; Ya-You et al. 1989). Other studies show that serum ferritin levels are significantly different in patients with benign and malignant effusions (abnormal fluid accumulations). Of the patients with benign nonneoplastic effusions, 96 percent had serum ferritin levels below 1,000 nanograms (ng) per ml; very high serum ferritin levels of over 3,000 ng per ml were encountered only in patients with malignancies (Yinnon et al. 1988).

Even more indicative is a study of patients afflicted with malignant histiocytosis and virus-associated hemophagocytic syndrome. At the onset of their disease these patients were anemic while having levels of serum ferritin above 1,000 ng per ml. As the disease progressed, serum ferritin levels increased to greater than 12,000 ng per ml (Esumi et al. 1988). Patients were given either chemotherapy or prednisone, depending on their disease. All patients with consistently high serum ferritin levels died within three months of treatment, whereas all patients with serum ferritin values less than 1,000 ng per ml lived beyond three months and are alive today with normal serum ferritin levels (Esumi et al. 1988). The more severe the disease, the higher the ferritin levels. We interpret these dramatic findings as the body's failed attempt to thwart disease by sequestering as much iron as possible. As the threat of disease decreased, serum ferritin levels began to drop. Noriko Esumi and colleagues (1988: 2071) conclude that "serum ferritin level in histiocytic proliferative disorders is a useful indicator of disease activity in both neoplastic and reactive conditions rather than only a marker of malignant process."

The popular belief that non-Western groups have a lower incidence of cancer because they have healthier diets than Westerners may be correct for reasons other than those usually proposed. As noted earlier,

Westerners tend to have high iron levels because of a generally high-iron diet, ubiquitous fortification of cereal products, and the widespread practice of taking vitamins fortified with iron. High iron levels have been implicated in increasing one's vulnerability to neoplasia. It might be that in an attempt to improve our diets to reduce morbidity through the ingestion of excess iron, we actually worsen our diets and encourage high levels of morbidity.

Hyperferremia: Causes and Problems

In addition to high serum ferritin values and high concentrations of iron in the liver and spleen, increased concentrations of iron and ferritin occur in the bile of hyperferremic individuals. Bile iron is increased in patients with idiopathic hemochromatosis to as much as twice that of normal individuals; bile ferritin is increased to as much as five times that of normal individuals (Hultcrantz et al. 1989).

There are basically two types of hyperferremia – acquired and inherited (Table IV.B.3.3). Inherited hyperferremia is usually termed hereditary (or idiopathic) hemochromatosis. This disorder is autosomal (involving a nonsex chromosome) recessive, the responsible gene being located on the short arm of chromosome 6 (Edwards et al. 1988: 1355). Although the full manifestation of the disease occurs only in affected homozygotes, a small proportion of heterozygotes have been found to exhibit minor abnormalities in laboratory tests that measure the body's iron burden (Edwards et al. 1988). The incidence of hemochromatosis in a presumably healthy population of 11,065 European-Americans (with transferrin saturation values of above 62 percent) was 5 people per 1,000; the amount of iron loading in the liver and consequent organ damage varied widely (Edwards et al. 1988). Studies conducted in Scotland and France yielded a lower number of affected individuals, although the incidence was still substantial, ranging from 1 in 400 to 1 in 517 persons (McLaren et al. 1983: 223).

Table IV.B.3.3. *Types of disorders associated with iron overload*

Primary iron overload

Hereditary (idiopathic) hemochromatosis

Secondary iron overload

Hereditary transferrin deficiency

Thalassemia syndromes

African siderosis (African hemosiderosis; also Bantu siderosis)

Kaschin-Beck disease

Transfusional hemosiderosis

Alcoholic cirrhosis with hemosiderosis

Porphyria cutanea tarda

Acquired hemolytic anemia

Ineffective erythropoiesis (red blood cell production)

Pyridoxine-responsive anemia

Source: Adapted from McLaren, Nuir, and Kellermeyer (1983: 206).

Thalassemia syndromes (i.e., thalassemia major, intermedia, minor, and minima) are inherited disorders in which intestinal iron absorption is increased due to the hemolytic anemia and ineffective erythropoiesis (i.e., production of red blood cells). The condition is complicated by transfusion therapy (McLaren et al. 1983: 216). The result is hyperferremia. Occasionally a complete absence of serum transferrin (termed atransferrinemia) occurs in which marrow is without iron but heavy iron deposits are found in the heart, pancreas, liver, and mucous glands (McLaren et al. 1983). The cause is linked to the homozygous expression of an autosomal recessive gene. Heterozygous expression of this recessive gene can also result in anemia.

More unusual is hereditary hypochromic anemia, which is a sex-linked recessive disorder that is variable in its manifestation. A poorly understood disorder, porphyria cutanea tarda, results from a defect in the heme synthetic pathway. This condition is often found in combination with hepatic siderosis (iron storage), possibly from increased dietary iron absorption or heterozygosity for idiopathic hemochromatosis (McLaren et al. 1983: 226).

Acquired hyperferremia results from a number of causes. Three of the most common disorders that produce acquired hyperferremia are transfusional iron overload (also called transfusional hemosiderosis), alcoholism and cirrhosis with hemosiderosis, and African siderosis (or Bantu siderosis). Iron overload can be a by-product of prolonged intravenous iron administration or repetitive transfusions. As a common complication of maintenance dialysis, for example, hyperferremia requires regular phlebotomy (bloodletting) to correct the situation (McCarthy et al. 1989).

Alcohol abuse stimulates absorption of dietary iron and reduces the ability of intestinal epithelial cells to prevent excessive transport of iron into the blood (Mazzanti et al. 1987). Hyperferremia can even occur in individuals who have not yet developed liver damage (Rodriguez et al. 1986).

In most parts of sub-Saharan Africa, African or Bantu siderosis results from brewing beer in iron containers and cooking food in cast-iron pots. Large deposits of iron occur in the liver, spleen, and to a lesser extent, in the macrophages of the bone marrow. One study indicated that the average rural southern African male ingests between 50 and 100 mg of iron daily in beer alone (Bothwell et al. 1965: 893). In a variety of studies of Africans who died from accidental deaths and other causes in South Africa, between 40 and 89 percent of the subjects exhibited varying degrees of iron overload (Bothwell and Bradlow 1960). The condition is detectable as early as late adolescence and becomes most severe between the ages of 40 and 60 years. In another study 75 percent of the males autopsied manifested siderosis compared with 25 percent of the women (Bothwell and Isaac-

son 1962: 524). The lower incidence of siderosis among women is due to a combination of greater iron loss through menstruation and the cultural practice of women drinking less beer and therefore ingesting less iron.

In the past 25 years the frequency of siderosis has dropped. This decline is particularly notable among urban Africans who do not drink as much of the traditional home-brewed beer as the rural population but still ingest large amounts of alcohol in different forms (MacPhail et al. 1979). A study conducted in the mid-1980s in Zimbabwe revealed that 12 percent of the men tested had an elevated serum ferritin level and a transferrin saturation of over 70 percent (Gordeuk, Boyd, and Brittenham 1986). However, although the frequency of siderosis in various areas of Africa has been declining over the past 25 to 30 years, it remains a serious health problem among a large percentage of rural populations.

Because hyperferremia results in the accumulation of iron in various organs, particularly the liver, spleen, heart, and pancreas, it can lead to severe cirrhosis, heart failure, and diabetes mellitus. Although the precise mechanism(s) involved is not well understood, it appears that tissue damage occurs in the liver and other organs as a result of the toxicity of excess iron. Cirrhosis occurs even in hyperferremic patients who do not drink any alcohol, although the combination of alcohol consumption and hyperferremia exacerbates the cirrhosis (McLaren et al. 1983: 240). In addition, different etiologies of iron overload appear to have slightly different manifestations. For example, the ratio of liver iron to marrow iron was much greater in patients with hereditary hemochromatosis than in those with African siderosis (McLaren et al. 1983: 236).

Hyperferremic individuals are plagued by problems other than those associated with the toxicity of excess iron. For reasons detailed earlier, they also suffer from an increased susceptibility to bacterial and parasitic microorganisms and neoplastic diseases (cancer). For example, one study demonstrated an increased risk of contracting *Yersinia enterocolitica* and *Yersinia pseudotuberculosis* bacteremia in dialysis patients with iron overload (Boelaert et al. 1987). Similar studies revealed an increased risk and virulence from several iron-sensitive microorganisms, including *Listeria monocytogenes*, *Brucella spp.*, and *Vibrio vulnificus*.

Problems Associated with Insufficient Iron

As in the case of hyperferremia, there are two basic categories of hypoferremia (or low circulating-iron levels) based on their etiology. Most common are the acquired anemias, including iron deficiency, drug-induced anemia, and anemia of chronic disease/inflammation. Less common are hereditary anemias, such as sickle-cell anemia and congenital

sideroblastic anemia. Some, such as sickle-cell anemia, are classified in a clinical setting by the morphological shape of the hemoglobin (Table IV.B.3.4).

Macrocytic anemia can be induced by diet, resulting from malabsorption, or it can be caused by an inherited disorder. Because Vitamin B₁₂ (cobalamin) and folic acid (folate) are required for normal red blood cell nuclear growth and synthesis, deficiencies of these nutrients can cause anemia (Simmons 1989: 12–14). Vitamin B₁₂ deficiency from dietary causes is very rare and occurs only in strict vegetarians who exclude all meat, eggs, and milk. However, this deficiency can also arise from a number of disorders, including impaired absorption of B₁₂ or folate such as in pernicious anemia; malabsorption that can result from certain drugs, such as those used to treat sprue and celiac diseases, and gastrectomy; competition from parasites such as the fish tapeworm; hereditary impairment of absorption capabilities; increased requirements of the vitamin due to pregnancy, tumors, and hyperthyroidism; or impaired utilization of the vitamin, as in red cell enzymopathy, abnormal binding proteins, absence of transport protein, or nitrous oxide administration (Simmons 1989: 40).

Folate deficiencies can result from the following: a lack of green vegetables in a diet; alcoholism; impaired absorption due to sprue and celiac diseases; drugs used to treat malignant diseases; malaria and bacterial infections; increased requirements stemming from pregnancy; conditions such as hyperthyroidism; or impaired utilization as occurs with drugs like phenytoin (Simmons 1989: 40–1).

Two types of hypochromic microcytic anemia are most prevalent. One is caused by blood loss that is not counterbalanced by a sufficient dietary intake. The second, less common cause results from unmet nutritional demands even without blood loss. In men, bleeding from the gastrointestinal tract is the most frequent cause, and although this condition may occur in women as well, menorrhagia (excessive menstrual flow) is more often responsible. The other principal cause of this variety of anemia is chronic disease, in which iron deficiency appears to form a nonspecific defense mechanism against disease.

Less common is sideroblastic anemia, which can be acquired or congenital. Acquired sideroblastic anemia can be drug- or toxin-induced from cancer chemotherapy, antituberculous drugs, and ethanol, or it can be idiopathic, such as preleukemic or dysmyelopoietic syndromes (Beissner and Trowbridge 1986). Patients with sideroblastic anemia tend to accumulate excess iron in the tissues and therefore are vulnerable to many of the problems associated with iron overload. Lead poisoning can also cause hypochromic microcytic anemia.

Some anemias are associated with specific geographical regions. For example, thalassemias are a varied group of inherited disorders characterized by one or more defects in the synthesis of the normal alpha or beta globin chains of hemoglobin. They can occur in homozygous or heterozygous states and include thalassemia major, thalassemia intermedia, thalassemia minor, and thalassemia minima (Simmons 1989: 55). The geographical distribution of thalassemia is primarily in the Mediterranean region, although it is also found in Southeast Asia, the Middle East, and the Orient and among immigrants from those areas. There is much variation in the clinical manifestations of these genetic disorders. The defect in hemoglobin chains causes a reduction in hemoglobin in afflicted individuals. Ironically, and as discussed earlier, thalassemia can also result in iron overload of some organs because intestinal iron absorption is increased due to the hemolytic anemia and ineffective production of red blood cells. In the heterozygous form, however, thalassemia has been postulated to be a deterrent to malaria infections, perhaps by causing a reduction in the amount of circulating iron.

A normocytic normochromic anemia that involves the premature destruction of red blood cells is sickle-cell anemia, an inherited autosomal dominant disorder. Sickle-cell anemia is lethal in the homozygous form, but it can also become symptomatic in heterozygotes in situations of oxygen deprivation such as at high altitudes. Geographically, sickle-cell anemia is most common in equatorial Africa but is also found to a lesser extent in the Mediterranean region and India. This distribution has been attributed to the improved immunity individuals who are heterozygous for the sickle-cell trait have from malaria parasites (e.g., *Plasmodium falciparum*). Approximately 8.5 percent of

Table IV.B.3.4. *Morphological classification of anemia*

Morphology	Anemia
Macrocytic	Nonmegaloblastic
	Liver disease
	Reticulocytosis
	Megaloblastic
	Vitamin B ₁₂ deficiency
	Pernicious anemia
	Folate deficiency
Hypochromic microcytic	Myelodysplastic syndrome
	Iron deficiency
	Anemia of chronic disease/inflammation
	Thalassemia
	Hereditary/acquired sideroblastic anemia
Normochromic normocytic	Lead poisoning
	Decreased reticulocytosis
	Renal failure
	Endocrinopathies
	Anemia of chronic disease/inflammation
	Aplasia (idiopathic or secondary)
	Myelophthisis
	Increased reticulocytosis
	Spherocytic (congenital and acquired)
	Nonspherocytic
	Sickle-cell anemia

American blacks are heterozygous for the sickle-cell trait and are relatively symptom-free; in contrast, homozygous individuals suffer from hemoglobin levels between 5 and 9 g per decaliter (dl), leg ulcers, hematuria, and other afflictions (Simmons 1989: 68–70).

Common autosomal dominant normocytic normochromic anemias that represent a defect in the red cell membrane include hereditary spherocytosis, found primarily in people of northern European descent, and hereditary elliptocytosis, found worldwide (Simmons 1989: 62–4). Spherocytic (congenital and acquired) anemia involves a deficiency in the Glucose-6-phosphate dehydrogenase enzyme (abbreviated G6PD) that is sex-linked with full expression in affected males and partial expression in females (Simmons 1989: 65). Its geographic distribution is worldwide, with the highest frequency in African populations, although it is also found in Italian, Greek, Asian, and Jewish populations.

Congenital nonspherocytic hemolytic anemia results from deficiency in several red cell enzymes, including glucose 6-phosphate dehydrogenase (G6PD). This condition is found in Asian, European, and Mediterranean populations, as well as in other populations to a lesser extent (Simmons 1989). This anemia is also caused by ingestion or contact with the fava bean by Mediterranean peoples (even from just inhaling the pollen of the bean among some sensitive males, usually under the age of 12). It can also be triggered by the use of antimalarial and other drugs in African and Mediterranean populations. A third cause can result from infection with viral or bacterial pathogens worldwide in persons with G6PD deficiency – pathogens include *Escherichia coli*, *Salmonella*, *Streptococcus*, *Rickettsiae*, viral hepatitis, pneumococcal pneumonia, malaria, and rheumatoid arthritis (Simmons 1989: 65).

A variety of anemias are associated with immunological disorders, including transfusion reactions, ABO or Rh blood-group incompatibility between fetus and mother, and autoimmune hemolytic anemia, a condition in which antibodies or lymphocytes attack cells of the person who produced them. The most common type of autoimmune hemolytic anemia is termed warm-antibody because the autoantibody reacts most efficiently with red cells at 37° Celsius (C) and occurs especially in systemic lupus erythematosus and lymphomas. Less common is cold-antibody type, in which the antibodies are optimally reactive at temperatures less than 37° C, including cold hemagglutinin disease and paroxysmal cold hemoglobinuria (Simmons 1989: 78–81).

Conclusions and Future Directions for Research

As we learn more about iron, we learn more of its multifaceted interrelationship with health and disease in a variety of conditions and situations. It is also apparent that this interrelationship is a complex one. For example, lowered iron levels do not necessarily

mean higher morbidity, nor is anemia primarily a nutritional disorder. We predict that future studies of iron will concentrate on its role in health, especially during the periods of rapid growth that occur in childhood and pregnancy, and in response to chronic infection, inflammation, and malignancy.

Further elucidation of iron as a cause of, and as a defense against, disease should reduce morbidity levels all over the world. Therefore, we have much to gain through multidisciplinary research that investigates all aspects of iron. Presented here is an overview of the fascinating picture of iron, a picture only now coming into focus.

Susan Kent

Patricia Stuart-Macadam

We are most grateful to Drs. Gene Weinberg, Steven Kent, and Roy Stuart for the valuable comments on a draft of this chapter. Any inadequacies, however, are solely our responsibility.

Notes

1. The reticuloendothelial system is also called the monocyte-macrophage system.
2. This excretion can be broken down into the following: gastrointestinal loss, of which blood comprises 0.35 mg, mucosal 0.10 mg, and biliary 0.20 mg; urinary loss of 0.08 mg; and skin loss of 0.20 mg (Fairbanks and Beutler 1988: 205).
3. Eugene Weinberg (1966, 1974, 1977, 1978, 1981, 1984, 1990) has numerous publications outlining the exact techniques microorganisms use to extract iron from hosts and the techniques hosts have developed to thwart them.
4. For example, most physicians routinely prescribe oral iron for their pregnant patients; others recommend iron supplements to patients with lowered hemoglobin levels without examining serum ferritin levels. We need to reassess the currently popular practice of indiscriminate iron fortification of foods and infant formula.

Bibliography

- Arthur, C. K., and J. P. Isbister. 1987. Iron deficiency: Misunderstood, misdiagnosed and mistreated. *Drugs* 33: 171–82.
- Behnke, J. M. 1987. Evasion of immunity by nematode parasites causing chronic infections. *Advances in Parasitology* 26: 1–71.
- Beissner, Robert, and Arthur Trowbridge. 1986. Clinical assessment of anemia. *Postgraduate Medicine* 80: 83–95.
- Bernat, Ivan. 1983. *Iron metabolism*. New York.
- Bezborovainy, Anatoly. 1980. *Biochemistry of nonheme iron in man*. New York.
- 1989a. Biochemistry of nonheme iron in man. I. Iron proteins and cellular iron metabolism. *Clinical Physiology and Biochemistry* 7: 1–17.
- 1989b. Biochemistry of nonheme iron in man. II. Absorption of iron. *Clinical Physiology and Biochemistry* 7: 53–69.
- Biemond, Peter, Antonius Swaak, Henk van Eijk, and Johan Koster. 1988. Superoxide dependent iron release from ferritin in inflammatory diseases. *Free Radical Biology and Medicine* 4: 185–98.

- Boelaert, Johan, Herman van Landuyt, Yvan Valcke, et al. 1987. The role of iron overload in *Yersinia enterocolitica* and *Yersinia pseudotuberculosis* bacteremia in hemodialysis patients. *Journal of Infectious Diseases* 156: 384–9.
- Bothwell, Thomas, R. D. Baynes, B. J. MacFarlane, A. P. MacPhail. 1989. Nutritional iron requirements and food iron absorption. *Journal of Internal Medicine* 226: 357–65.
- Bothwell, Thomas, and Basil Bradlow. 1960. Siderosis in the Bantu. *Archives of Pathology* 70: 279–92.
- Bothwell, Thomas, Robert Charlton, James Cook, and Clement Finch. 1965. Oral iron overload. *South African Medical Journal* 30: 892–900.
1979. *Iron metabolism in man*. Oxford.
- Bothwell, Thomas, and C. Isaacson. 1962. Siderosis in the Bantu: A comparison of incidence in males and females. *British Medical Journal* 1: 522–4.
- Bullen, J. J., and E. Griffiths. 1987. *Iron and infection*. London.
- Bullen, J. J., H. Rogers, and E. Griffiths. 1974. Bacterial iron metabolism in infection and immunity. In *Microbial iron metabolism: A comprehensive treatise*, ed. J. B. Neiland, 383–439. New York.
- Burns, Edward, Nahum Goldberg, Christine Lawrence, and Barry Wenz. 1990. Clinical utility of serum tests for iron deficiency in hospitalized patients. *American Journal of Clinical Pathology* 93: 240–5.
- Cerioti, Ferruccio, and Giovanni Cerioti. 1980. Improved direct specific determination of serum iron and total iron-binding capacity. *Clinical Chemistry* 26: 327–31.
- Colditz, G. A., W. C. Willett, M. J. Stampfer, et al. 1987. Menopause and the risk of coronary heart diseases in women. *New England Journal of Medicine* 316: 1105–10.
- Cook, James. 1990. Adaptation in iron metabolism. *American Journal of Clinical Nutrition* 51: 301–8.
- Cook, James, and Barry Skikne. 1982. Serum ferritin: A possible model for the assessment of nutrient stores. *American Journal of Clinical Nutrition* 35: 1180–5.
1989. Iron deficiency: Definition and diagnosis. *Journal of Internal Medicine* 226: 349–55.
- Crosa, J. H. 1987. Bacterial iron metabolism, plasmids and other virulence factors. In *Iron and infection*, ed. J. J. Bullen and E. Griffiths, 139–70. London.
- Dallman, P. R. 1989. Iron deficiency: Does it matter? *Journal of Internal Medicine* 226: 367–72.
- Dunn, Frederick. 1968. Epidemiological factors: Health and disease in hunter-gatherers. In *Man the hunter*, ed. Richard Lee and Irvn De Vore, 221–8. Chicago.
- Eaton, S. B., M. Shostak, and M. Konner. 1988. *The paleolithic prescription*. New York.
- Edwards, Corwin, Linda Griffen, David Goldgar, et al. 1988. Prevalence of hemochromatosis among 11,065 presumably healthy blood donors. *New England Journal of Medicine* 318: 1355–62.
- Elin, Ronald, Sheldon Wolff, and Clement Finch. 1977. Effect of induced fever on serum iron and ferritin concentrations in man. *Blood* 49: 147–53.
- Esumi, Noriko, Satoshi Ikushima, Shigeyoshi Hibi, et al. 1988. High serum ferritin level as a marker of malignant histiocytosis and virus-associated hemophagocytic syndrome. *Cancer* 61: 2071–6.
- Fairbanks, Virgil, and Ernest Beutler. 1988. Iron. In *Modern nutrition in health and disease*, ed. Maurice Shils and Vernon Young, 193–226. Philadelphia, Pa.
- Farley, Patrick, and Jaime Foland. 1990. Iron deficiency anemia: How to diagnose and correct. *Postgraduate Medicine* 87: 89–101.
- Fielding, Jack. 1980. Serum iron and iron binding capacity. In *Iron*, ed. James Cook, 15–43. New York.
- Finch, C. A. 1989. Introduction: Knights of the oval table. *Journal of Internal Medicine* 226: 345–8.
- Finch, C. A., and H. Huebers. 1982. Perspectives in iron metabolism. *New England Journal of Medicine* 306: 1520–8.
- Florentino, Rodolfo, and Romualda Guirriec. 1984. Prevalence of nutritional anemia in infancy and childhood with emphasis on developing countries. In *Iron nutrition in infancy and childhood*, ed. A. Stekel, 61–74. New York.
- Frassinelli-Gunderson, E. P., S. Margen, and J. R. Brown. 1985. Iron stores in users of oral contraceptive agents. *American Journal of Clinical Nutrition* 41: 703–12.
- Goodner, Kenneth. 1933. Bacteriological and serological studies in Yucatan in 1929. In *The peninsula of Yucatan: Medical, biological, meteorological and sociological studies*, ed. George Shattuck, 201–20. Carnegie Institution of Washington Publication No. 431.
- Gordeuk, Victor, R. Devee Boyd, and Gary Brittenham. 1986. Dietary iron overload persists in rural sub-Saharan Africa. *Lancet* 1310–13.
- Griffiths, E. 1987. Iron in biological systems. In *Iron and infection*, ed. J. J. Bullen and E. Griffiths, 1–26. London.
- Griffiths, E., and J. J. Bullen. 1987. Iron-binding proteins and host defence. In *Iron and infection*, ed. J. J. Bullen and E. Griffiths, 171–210. London.
- Guyatt, Gordon, Christopher Patterson, Mahomoud Ali, et al. 1990. Diagnosis of iron-deficiency anemia in the elderly. *American Journal of Medicine* 88: 205–9.
- Hallberg, Leif. 1980. Food iron absorption. In *Iron*, ed. James Cook, 117–33. New York.
- Hann, Hie-Won, Young Chung, W. Thomas, and Baruch Blumberg. 1989. Increased serum ferritin in chronic liver disease: A risk factor for primary hepatocellular carcinoma. *International Journal of Cancer* 43: 376–9.
- Herold, P. M., and J. E. Kinsella. 1986. Fish oil consumption and decreased risk of cardiovascular disease. *American Journal of Clinical Nutrition* 43: 566–98.
- Hoffbrand, A. V., and S. M. Lewis. 1981. *Postgraduate haematology*. London.
- Houwelingen, R. V., A. Nordoy, E. van der Beek, et al. 1987. Effect of a moderate fish intake on blood pressure, bleeding time, hematology, and clinical chemistry in healthy males. *American Journal of Clinical Nutrition* 46: 424–36.
- Hultcrantz, Rolf, Bo Angelin, Erik Björn-Rasmussen, et al. 1989. Eiliary excretion of iron and ferritin in idiopathic hemochromatosis. *Gastroenterology* 96: 1539–45.
- Jacobs, A., and M. Worwood. 1982. Iron metabolism, iron deficiency and overload. In *Blood and its disorders*. Second edition, ed. R. M. Hardisty and D. J. Weatherall, 149–97. Oxford.
- Kent, Susan. 1986. The influence of sedentism and aggregation on porotic hyperostosis and anaemia: A case study. *Man* 21: 605–36.
1992. Anemia through the ages. In *Diet, demography and disease: Changing perspectives of anemia*, ed. Patricia Stuart-Macadam and Susan Kent, 1–30. New York.
- Kent, Susan, and Eugene Weinberg. 1989. Hypoferremia: Adaption to disease? [letter] *New England Journal of Medicine*: 320, 672.
- Kent, Susan, Eugene Weinberg, and Patricia Stuart-Macadam. 1990. Dietary and prophylactic iron supplements: Helpful or harmful? *Human Nature* 1: 53–79.
1994. The etiology of the anemia of chronic disease. *Journal of Clinical Epidemiology* 47: 23–33.

- Keusch G., and M. Farthing. 1986. Nutrition and infection. *Annual Review of Nutrition* 6: 131-54.
- Kluger, M. J., and J. J. Bullen. 1987. Clinical and physiological aspects. In *Iron and infection*, ed. J. J. Bullen and E. Griffiths, 243-82. London.
- Kochan, I. 1973. The role of iron in bacterial infections, with special consideration of host-tubercle bacillus interaction. *Current Topics in Microbiology and Immunology* 60: 1-30.
- Kromhout D., E. B. Bosschieter, and C. de Lezenne Coulander. 1985. The inverse relation between fish consumption and 20-year mortality from coronary heart disease. *New England Journal of Medicine* 312: 1205-9.
- Labbe, Robert, and Clement Finch. 1980. Erythrocyte protoporphyrin: Application of the diagnosis of iron deficiency. In *Iron*, ed. James Cook, 44-58. New York.
- Lee, G. R. 1983. The anemia of chronic disease. *Seminar in Hematology* 20: 61-80.
- Leon, M. B., J. S. Borer, S. L. Bacharach, et al. 1979. Detection of early cardiac dysfunction in patients with severe beta-thalassemia and chronic iron overload. *New England Journal of Medicine* 301: 1143-8.
- Liebel, R. L., A. B. Greenfield, and E. Pollitt. 1979. Iron deficiency: Behavior and brain chemistry. In *Nutrition: Pre- and postnatal development*, ed. Myron Winick, 383-439. New York.
- Long, P. J., and B. Shannon. 1983. *Focus on nutrition*. Englewood Cliffs, N. J.
- MacPhail, A. P., M. O. Simon, J. D. Torrance, et al. 1979. Changing patterns of dietary iron overload in Black South Africans. *American Journal of Clinical Nutrition* 32: 1272-8.
- Martinez, J. L., A. Delgado-Irabarren, and F. Baquero. 1990. Mechanisms of iron acquisition and bacterial virulence. *FEMS: Microbiology Reviews* 75: 45-56.
- Masawe, A. E., J. M. Muindi, and G. B. Swai. 1974. Infections in iron deficiency and other types of anaemia in the tropics. *Lancet* 2: 314-17.
- McCarthy, James, William Johnson, David Nixon, et al. 1989. Transfusional iron overload in patients undergoing dialysis. *Journal of Laboratory and Clinical Medicine* 114: 193-9.
- McLaren, Gordon, W. Angus Muir, and Robert Kellermeyer. 1983. Iron overload disorders: Natural history, pathogenesis, diagnosis, and therapy. *Critical Reviews in Clinical Laboratory Sciences* 19: 205-66.
- Mazzanti, R., K. S. Srai, E. S. Debnam, et al. 1987. The effect of chronic ethanol consumption on iron absorption in rats. *Alcohol and Alcoholism* 22: 47-52.
- Miller, Carole, Richard Jones, Steven Plantadosi, et al. 1990. Decreased erythropoietin response in patients with the anemia of cancer. *New England Journal of Medicine* 322: 1689-92.
- Monsen, Elaine. 1988. Iron nutrition and absorption: Dietary factors which impact iron bioavailability. *Journal of the American Dietetic Association* 88: 786-90.
- Moore, Carl. 1973. Iron. In *Modern nutrition in health and disease*. Fifth edition, ed. R. S. Goodhart and M. E. Shils. Philadelphia, Pa.
- Murray M. J., and A. Murray. 1977. Starvation suppression and refeeding activation of infection: An ecological necessity? *Lancet* 1: 123-5.
- Murray M. J., A. Murray, and C. J. Murray. 1980. An ecological interdependence of diet and disease? *American Journal of Clinical Nutrition* 33: 697-701.
- Murray, M. J., A. Murray, N. Murray, and M. Murray. 1976. Somali food shelters in the Ogaden famine and their impact on health. *Lancet* 1: 1283-5.
- Noyes, Ward. 1985. Anemia as a result of the insufficiency in the production of red cells. In *Hematology and oncology*, ed. Marshall Lichtman, 23-8. New York.
- Oppenheimer, S. J., F. D. Gibson, S. B. MacFarlane, et al. 1986. Iron supplementation increases prevalence and effects of malaria: Report on clinical studies in Papua New Guinea. *Transactions of the Royal Society of Tropical Medicine and Hygiene* 80: 603-12.
- Osborne, Pamela, Luther Burkett, George Ryan, and Marlene Lane. 1989. An evaluation of red blood cell heterogeneity (increased red blood cell distribution width) in iron deficiency of pregnancy. *American Journal of Obstetrics and Gynecology* 160: 336-9.
- Payne, S. 1988. Iron and virulence in the family Enterobacteriaceae. *CRC Critical Reviews in Microbiology* 16(2): 81.
- Pearson, H., and J. Robinson. 1976. The role of iron in host resistance. *Advanced Pediatrics* 23: 1-33.
- Pochedly, C., and S. L. May. 1987. Iron deficiency anemia in children. *American Family Physician* 35: 195-200.
- Picciano, M. F., and H. A. Guthrie. 1976. Copper, iron and zinc contents of mature human milk. *American Journal of Clinical Nutrition* 29: 242-54.
- Reizenstein, Peter. 1983. *Hematological stress syndrome: The biological response to disease*. New York.
- Rodriguez, M. C., M. S. Henriquez, A. F. Turon, et al. 1986. Trace elements in chronic alcoholism. *Trace Elements in Medicine* 3: 164-7.
- Salonen, J. T., K. Nyyssönen, H. Korpela, et al. 1992. High stored iron levels are associated with excess risk of myocardial infarction in Finnish men. *Circulation* 86: 803-11.
- Schade, A. L., and L. Caroline. 1944. Raw hen egg white and the role of iron in growth inhibition of *Shigella dysenteriae*, *Staphylococcus aureus*, *Escherichia coli*, and *Saccharomyces cerevisiae*. *Science* 100: 14-15.
- Selby, Joseph, and Gary Friedman. 1988. Epidemiologic evidence of an association between body iron stores and risk of cancer. *International Journal of Cancer* 41: 677-82.
- Simmons, Arthur. 1989. *Hematology: A combined theoretical and technical approach*. Philadelphia, Pa.
- Stevens, R. G., D. Y. Jones, M. S. Micozzi, and P. R. Taylor. 1988. Body iron stores and the risk of cancer. *New England Journal of Medicine* 319: 1047-52.
- Strauss, R. 1978. Iron deficiency, infections, and immune function: A reassessment. *American Journal of Clinical Nutrition* 31: 660-6.
- Stuart-Macadam, Patricia. 1988. Nutrition and anemia in past human populations. In *Diet and subsistence: Current archaeological perspectives*, ed. B. Kennedy and G. LeMoine, 284-7. Calgary.
- Subar, Amy, and Jean Bowering. 1988. The contribution of enrichment and fortification to nutrient intake of women. *Journal of the American Dietetic Association* 88: 1237-45.
- Sullivan, J. L. 1983. The sex difference in ischemic heart disease. *Perspectives in Biology and Medicine* 26: 657-71.
1989. The iron paradigm of ischemic heart disease. *American Heart Journal* 117: 1177-88.
- Thompson, Warren. 1988. Comparison of tests for diagnosis of iron depletion in pregnancy. *American Journal of Obstetrics and Gynecology* 159: 1132-4.
- Thompson, Warren, Thomas Meola, Mack Lipkin, and Michael Freedman. 1988. Red cell distribution width, mean corpuscular volume, and transferrin saturation in the diagnosis of iron deficiency. *Archives of Internal Medicine* 148: 2128-30.

- Wadsworth, George. 1975. Nutritional factors in anemia. *World Review of Nutrition and Diet* 21: 75–150.
1991. Iron and anemia. Paper presented at the American Association of Physical Anthropology Conference, Milwaukee, Wis.
- Weinberg, Eugene. 1966. Roles of metallic ions in host-parasite interactions. *Microbiological Reviews* 30: 136–51.
1974. Iron and susceptibility to infectious disease. *Science* 184: 952–6.
1977. Infection and iron metabolism. *American Journal of Clinical Nutrition* 30: 1485–90.
1978. Iron and infection. *Microbiological Review* 42: 45–66.
1981. Iron and neoplasia. *Biology and Trace Element Research* 3: 55–80.
1983. Iron in neoplastic disease. *Nutrition and Cancer* 4: 223–33.
1984. Iron withholding: A defense against infection and neoplasia. *Physiological Reviews* 64: 65–102.
1989. Cellular regulation of iron assimilation. *Quarterly Review of Biology* 64: 261–90.
1990. Cellular iron metabolism in health and disease. *Drug and Metabolism Reviews* 22: 531–79.
1992. Iron withholding: A defense against disease. In *Diet, demography, and disease: Changing perspectives of anemia*, ed. Patricia Stuart-Macadam and Susan Kent. New York.
- Woods, S., T. DeMarco, and M. Friedland. 1990. Iron metabolism. *American Journal of Gastroenterology* 85: 1–8.
- Worwood, M. 1980. Serum ferritin. In *Iron in biochemistry and medicine, II*, ed. A. Jacobs and M. Worwood, 203–44. London.
1989. An overview of iron metabolism at a molecular level. *Journal of Internal Medicine* 226: 381–91.
- Ya-You, Ji, Liu Yan-Fang, Wang Bo-Yun, and Yang De-Yun. 1989. An immunocytochemical study on the distribution of ferritin and other markers in 36 cases of malignant histiocytosis. *Cancer* 64: 1281–9.
- Yinnon, Amos, Abraham Konijn, Gabriela Link, et al. 1988. Diagnostic value of ferritin in malignant pleural and peritoneal effusions. *Cancer* 62: 2564–8.
- Zanella, Alberto, Linda Gridelli, Alessandra Berzuini, et al. 1989. Sensitivity and predictive value of serum ferritin and free erythrocyte protoporphyrin for iron deficiency. *Journal of Laboratory and Clinical Medicine* 113: 73–8.

IV.B.4 Magnesium

*It's astonishing how much you may see,
in a thicker fog than that, if you only
take the trouble to look for it.*

Charles Dickens, *Christmas Stories*

Magnesium is one of the most plentiful elements in nature and the fourth most abundant metal in living organisms. It is extremely important in both plant and animal metabolism. Photosynthesis does not proceed when the magnesium atom is removed from the chlorophyll molecule. Magnesium also plays a key role in many enzyme reactions that are critical to cellular metabolism and is one of the main determinants of biological excitation (Aikawa 1981).

Despite its ubiquitous distribution and the multiplicity of its actions, magnesium has long been considered as a microelement with a vague physiological role, and not until the early 1930s was it recognized as an essential nutrient. Magnesium deficiency in humans was only described in 1951, and according to several experts, it continues to be diagnosed less frequently than it should be (Whang 1987).

There are many explanations for the reluctance to allow magnesium deficiency a place in medicine; among them are the difficulties in measuring magnesium, which have restrained the accumulation of knowledge. Moreover, the essentially intracellular location of the magnesium ion has discouraged the detection of its deficit. Lastly, because magnesium is so widely distributed in foods, its dietary intake has been assumed to be sufficient to meet the body's requirements.

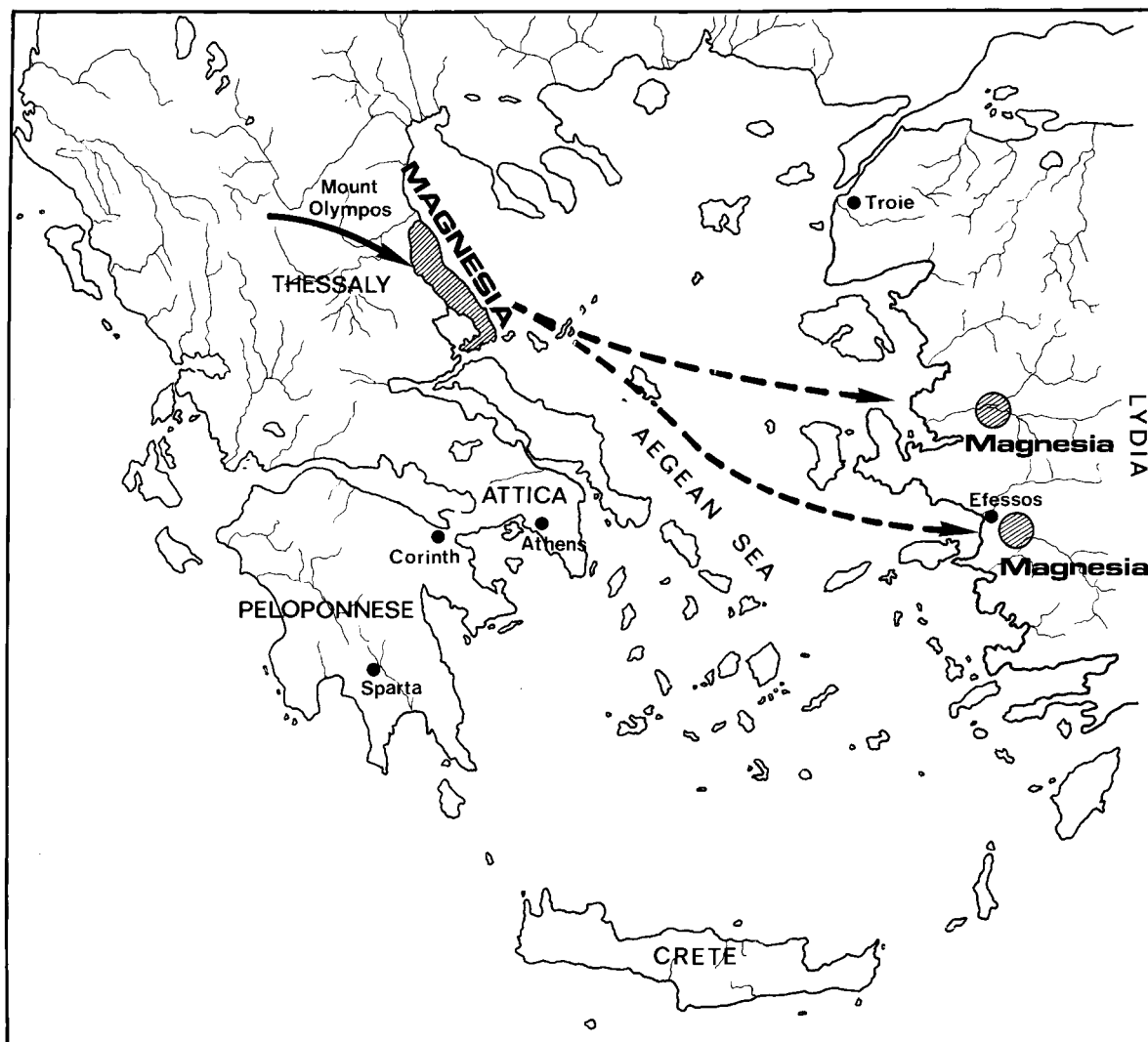
Actually, pure magnesium deficiency is quite rare. Marginal deficiency of magnesium, however, is believed to occur fairly often in the general population. Moreover, in most diseases causing magnesium deficiency, significant nutritional factors exist.

A Note on Etymology

The word “magnesium” originates from the Greek word *Magnesia*. Magnesia is the eastern peninsular region of Thessaly in central Greece. It was named after the *Magnetes*, a prehistoric Macedonian people, by whom Magnesia was first inhabited in the twelfth century B.C. The *Magnetes* are mentioned by Homer, and the port of Magnesia in the Pagassitic Gulf was the place whence the Argonauts sailed on their way to Colchide. Magnesia was also the name of three different overseas colonies established by the *Magnetes* during the eleventh century B.C.: one in central Crete, and two in the inland of Asia Minor (Map IV.B.4.1).

Magnetes lithos and *Magnesia lithos*, which both mean “stone of Magnesia,” were the names given by the ancient Greeks to minerals obtained from the earth of either the metropolitan Magnesia or its Asian colonies. *Magnetes lithos* or *magnetes* (magnet) was the term used for lodestone (native iron oxide that attracts iron). Hippocrates, however, the fifth-century B.C. father of medicine, referred to *Magnesia lithos* as a cathartic (purgative). He recommended (as translated by Paul Potter) that “if the cavity does not have a spontaneous movement, you must clean it out by giving spurge-flax, Cnidian berry, hippopheos, or magnetic stone” (Hippocrates 1988). (“Stone of Magnesia” or “Magnesian stone” would be a more appropriate translation than “magnetic stone.”) Presumably, the stone mentioned by Hippocrates was native magnesium carbonate (magnesite), the white and light powder that was later called *magnesia alba* and is still used as an aperient and antacid.

Magnesia alba was prepared from the mother liquors obtained in the manufactures of niter by M. B. Valentini of Giessen, in 1707, but it was confused with “calcareous earth” until 1755, when J. Black of Edin-



Map IV.B.4.1. Magnesia and its colonies in Asia Minor. The migration of *Magnetes* during the twelfth and eleventh centuries B.C.

burgh distinguished chemical differences between magnesia and lime. In 1807, Sir Humphrey Davy, conducting his studies on earth compounds in London, succeeded in isolating a number of alkali-earth metals that he named barium, strontium, calcium, and magnesium after their oxides baryta, strontia, chalk, and magnesia. Before long, "magnesium" replaced magnesium as the name of the element derived from magnesia (Durlach 1988b).

Recognition of Magnesium Deficiency

The Early Years

It was near the end of the nineteenth century when the nutritional significance of minerals was first recognized. That not only organic substances but also nonor-

ganic elements contained in food are necessary for life became evident in about 1870, when the German biologist J. Forster demonstrated the lethal effect in dogs of a diet deprived of inorganic salts (Forster 1873). A few years later, J. Gaube du Gers in France reported that mice fed a diet consisting of magnesium-free bread and distilled water were rendered progressively sterile. Impressed by this finding, he hastened to conclude that magnesium is "the metal of vital activity for what is most precious in life: reproduction and sensation" (Gaube 1895).

By 1900, on the American side of the Atlantic, J. Loeb accomplished his original experiments on the physiology of neuromuscular contractions and was able to state that biological excitation "depends upon the various ions, especially the metal ions

(sodium, calcium, potassium, and magnesium) existing in definite proportions in the tissue” (Loeb 1900) (Table IV.B.4.1). About the same time, it was discovered that the magnesium atom occupies the central position in the chlorophyll molecule (Willstatter and Stoll 1913). Thus the most fundamental role of magnesium in nature was discovered, for which a 1915 Nobel Prize was awarded.

As postulated by Jerry Aikawa in recent years, life on earth would not be possible without magnesium, since photosynthesis is absolutely dependent on this element and the photosynthetic process is indispensable for the enrichment of the environment with oxygen (Aikawa 1981). An understanding, however, of the myriad roles of magnesium in biology had to await a method suitable for its measurement in biological materials. The first such method was a gravimetric technique involving the precipitation of magnesium ammonium phosphate. It was introduced by L. B. Mendel and S. R. Benedict in 1909 and was subsequently superseded by its modification to a colorimetric procedure (Alcock 1969).

In the following years, a large number of chemical approaches, including adsorption of titan yellow dye by magnesium hydroxide in the presence of a stabilizer, titrimetric, or colorimetric methods measuring various magnesium-dye complexes, and fluorometric techniques (Table IV.B.4.2), were used for the determination of magnesium in plasma or serum, urine, or fecal samples. The multiplicity of methods is indicative of the lack of satisfaction among investigators with any one of them.

An impressive step forward in the understanding of the role of magnesium metabolism in health and disease was made by the availability of atomic absorption spectrophotometry. This simple, precise method permits multiple determinations of minute amounts of magnesium in small samples of biological fluids

Table IV.B.4.2. *Year of first application of different procedures for measuring magnesium in biological materials*

1909	Magnesium ammonium phosphate precipitation
1927	Titan yellow method
1951	EDTA titration
1952	Emission (flame) spectrophotometry
1955	Colorimetric procedures
1959	Fluorometric techniques
1960	Atomic absorption spectrophotometry

and tissues, in a matter of minutes. The theoretical basis of flame spectrophotometry lies in the discovery made by the famous German physicist Gustav Kirchhoff in 1860, that both emission and absorption of light of a specific wavelength are characteristic of a given element. Emission spectrophotometry was first used to measure serum magnesium by V. Kapuscinski and his co-workers in 1952. The quantitative measurement of atomic absorption was conceived by A. Walsh in 1955. J. Alan applied this technique to the measurement of magnesium in plants in 1958, and two years later, atomic absorption spectrophotometry was used for the determination of serum magnesium by J. Willis (Alcock 1969).

Alkaline phosphatase was the first enzyme shown to be activated by magnesium. The discovery was made by H. Erdtmann in 1927. Since then, as many as 300 enzymes have been known to be activated by this cation *in vitro*, including all those utilizing adenosine triphosphate or catalyzing the transfer of phosphate. Because magnesium is the second most plentiful intracellular cation, it has been postulated that its function also extends to all these enzymes *in vivo*. By inference, the predominant view has been that magnesium is required for most of the major metabolic pathways in the cell, including membrane transport, protein, nucleic acid, fat and coenzyme synthesis, glucose utilization, and oxidative phosphorylation (Wacker 1969).

Introducing the Concept

The credit for introducing the concept of magnesium deficiency belongs to Jehan Leroy. He showed for the first time that the metal is essential for mammalian metabolisms when he found that white mice fed a diet deficient in magnesium failed to grow (Leroy 1926). However, the most distinctive manifestation of magnesium deficiency was recognized in 1932, when the group led by E. McCollum reported the development of hyperemia and progressive neuromuscular irritability, culminating in generalized and sometimes fatal seizures, in weanling rats fed a diet containing 0.045 millimoles (mmol) per kilogram (kg) of the element (Kruse, Orent, and McCollum 1932).

It should be noted that flaccid paralysis as a pharmacological effect of intravenous magnesium sulfate has been described by French authors since 1869 (Jolyet and Cahours 1869). A distinction, however, should be made between pharmacological properties

Table IV.B.4.1. *Some important dates*

1807	Discovery of magnesium by Sir Humphrey Davy
1900	Experimental evidence that irritability of the nervous system depends upon metal ions, including magnesium (J. Loeb)
1903	Discovery that the magnesium atom occupies the central location in the chlorophyll molecule (R. Willstatter)
1909	Development of a chemical method for the determination of magnesium in biological material (L. Mendel and S. Benedict)
1926	Demonstration that magnesium is essential for growth and life (J. Leroy)
1932	Attribution of grass tetany of ruminants to hypomagnesemia (B. Sjollema)
1934	Description of a clinical syndrome of hypomagnesemia associated with twitching or convulsions (A. Hirschfelder and V. Haury)
1951	Description of clinical magnesium deficiency (E. Flink)
1960	Measurement of serum magnesium by atomic absorption spectrophotometry (J. Willis)

of a substance and manifestations of its metabolic effect. The fact that a generalized convulsion can be controlled by massive parenteral magnesium administration, for instance, does not necessarily mean that the symptom is due to deficiency of this ion.

During the 1950s and 1960s, the early experiments of McCollum and his colleagues were confirmed and extended by many research groups. Experimental magnesium deficiency produced in rats, dogs, cocks, and other animal species was found to be associated with multiple pathological lesions, including nephrocalcinosis, myocardial necrosis, and calcium deposition in the aorta and the coronary and peripheral vessels; and to be accompanied by other metabolic changes, such as hypercalcemia, azotemia, and tissue potassium depletion.

About the time that McCollum published his original findings on experimental magnesium deficiency, a form of spontaneously occurring hypomagnesemic tetany was observed in adult lactating cattle and sheep. The disease, characterized by neuromuscular irritability, tetany, and convulsions, was termed "grass tetany" because the symptoms developed in the affected animals when they were first allowed to graze on fresh green grass in the spring (Sjollem 1932). Grass tetany has since been reported in many parts of the world, including the United States, many European countries, New Zealand, Australia, and Japan. Although there have been numerous investigations, the etiology of the disease is not clear.

Plasma magnesium is usually low in those afflicted, but grass tetany has also been observed in animals with decreased cerebrospinal magnesium levels whose serum magnesium was normal. The diet of the animals, when they are confined indoors during the winter, consists of silage, grains, and grain mash, which are deficient in magnesium, but symptoms of deficiency do not arise until shortly – sometimes within two days – after the animals are turned out on spring grass. It is of interest that the magnesium content of grasses and grazing vegetation in grass tetany-prone areas has been found to be decreased when compared with the vegetation of non-prone areas (Kubota 1981).

In 1959, a second form of hypomagnesemic tetany of ruminants was recognized when R. Smith reported a gradual development of hypomagnesemia in association with progressive irritability and tetany in calves fed for a long time on a pure milk diet. Magnesium supplementation was found to prevent the disease (Smith 1959). It should be noted that cow's milk contains magnesium in fair amounts, but its high phosphate and calcium content can adversely interfere with magnesium absorption.

Magnesium Deficiency in Humans

As early as 1932, B. Sjollem and L. Seekles, by extrapolating data from veterinary medicine, postulated that human cases of tetany may be related to hypomagne-

semia. In 1934, A. B. Hirschfelder and V. G. Haury described seven patients with hypomagnesemia associated with muscular twitching or convulsions and concluded that a clinical syndrome did indeed exist involving low magnesium (hypomagnesemia) accompanied by twitching or by convulsions. One more case of hypomagnesemic tetany was reported by J. E. Miller in 1944. The patient was a 6-year-old boy with associated osteochondrosis of the capital epiphysis of the femurs. Soon after World War II inanition was recognized as a cause of hypomagnesemia, but clinical manifestations were not mentioned (Flink 1980).

The broad spectrum of full-blown manifestations of the magnesium deficiency syndrome was recognized in the early 1950s, when Edmund B. Flink described the amazing case of a woman who, having been on almost continuous intravenous fluid therapy for several months, developed striking neurological symptoms and signs, including almost every form of involuntary movement (Table IV.B.4.3). When first seen by Flink in July 1951 at Minnesota University Hospital, the patient was cachectic, dehydrated, and semicomatose and had severe hyponatremia, hypokalemia, hypophosphatemia, and hypochloremic alkalosis. With the appropriate replacement treatment, serum electrolytes returned to normal within a few days. However, on the sixth day of treatment, the patient began having repeated convulsions associated with gross tremor and myoclonic jerks of extremities, jaw, and tongue, facial

Table IV.B.4.3. *Causes of human magnesium deficiency and year of their first description*

Causes	Year
Inadequate intake	
Starvation	1949
Prolonged fluid therapy without magnesium	1951
Chronic alcoholism	1954
Protein-caloric malnutrition	1960
Defective absorption	
Prolonged diarrhea - any cause	1957
Malabsorption	1960
Increased urinary losses	
Diuretic-induced	1952
Renal diseases	1958
Drug-induced tubular dysfunction	1969
Endocrine and metabolic disorders	
Uncontrolled diabetes with ketoacidosis	1947
Hyperthyroidism	1955
Aldosteronism	1955
Hyperparathyroidism	1956
Other	
Neonatal hypocalcemia, hypomagnesemia	1962
Excessive lactation	1963
Genetic	1965

Source: Data from Wacker and Parisi (1968), Flink (1980), Whang (1987), and Durlach (1988a).

grimacing, choreiform and athetoid movements, and inability to talk or swallow.

These symptoms continued until the eleventh day, when a blood sample was taken for the determination of serum magnesium level. As stated quite emphatically by Flink, "serum magnesium level was obtained for no better reason than that magnesium and phosphate interact and are important intracellular elements" (Flink 1985). Serum magnesium was determined by the titan yellow method and was found to be very low. This finding prompted the intramuscular administration of magnesium sulfate in the daily dose of 2.0 mg. The response was so dramatic that it was characterized by Flink as "unforgettable" (Flink 1985). Tremor decreased after the first dose, and within 24 hours the patient was oriented and able to speak and eat (Fraser and Flink 1951).

One year later, similar symptoms in a patient with alcoholic cirrhosis stimulated the start of clinical and laboratory research on magnesium metabolism in alcoholism by Flink and his colleagues. Much of the knowledge of normal human magnesium metabolism and its alteration in disease originated from these early studies (Wacker and Parisi 1968).

Initially, the discovery of hypomagnesemia in patients with chronic alcoholism and delirium tremens, together with the response of some of these patients to parenteral administration of magnesium, led to the hypothesis that symptoms of alcohol withdrawal were due to magnesium deficiency. However, it was soon realized that many alcoholic patients with near normal serum magnesium levels also had symptoms, whereas others with very low levels did not. It was, therefore, suggested that serum magnesium might not reflect magnesium status, and balance studies were undertaken to elucidate this issue further. It is now quite clear that the main cause of magnesium depletion in alcoholism is inadequate magnesium intake. These early balance studies established the concept of tissue magnesium depletion in the presence of normal serum magnesium levels and revealed increased urinary losses as a possible mechanism of magnesium deficiency. Thus, a significant amount of parenterally administered magnesium was found to be retained by alcoholic patients, indicating tissue magnesium depletion, whereas acute administration of large doses of alcohol was shown to result in increased renal magnesium excretion (Flink et al. 1954). Urinary excretion of magnesium after a load dose has been subsequently established as a feasible test in clinical practice for the evaluation of magnesium status.

That magnesium depletion is likely to occur as a result of drug treatment was first recognized in 1952, when H. E. Martin, J. Mehl, and M. Wertman reported a fall in serum magnesium in association with increased magnesium excretion in 10 patients with congestive heart failure treated with ammonium chloride and mercurial diuretics. Since that time, other diuretics,

such as thiazides and loop diuretics; antibiotics, including aminoglycosides, tetracycline, carbenicillin, and amphotericin B; and also cisplatin and cyclosporin have been added to the list of iatrogenic causes of magnesium depletion (Whang 1987).

Intestinal disorders were recognized quite early as a common cause of clinical magnesium deficiency. During the late 1950s and the early 1960s, ulcerative colitis and regional enteritis, along with almost every cause of persistent diarrhea, including chronic laxative abuse, and also intestinal malabsorption due to chronic pancreatitis, short bowel syndrome, gluten enteropathy and tropical sprue, have been reported fairly often in association with symptomatic hypomagnesemia (Table IV.B.4.3).

Data from patients with malabsorption have confirmed previous observations in patients with primary hyperparathyroidism, indicating a relationship between magnesium metabolism and calcium metabolism. Hyperparathyroidism has been recognized as a cause of hypomagnesemia since 1956 (Harman 1956). Although some patients have been found to develop symptoms while hypercalcemic, in most cases hypomagnesemia has accompanied the fall in serum calcium levels after the removal of a parathyroid adenoma and has been attributed to the absorption of magnesium by the "hungry" bones of osteitis fibrosa cystica (Barnes, Krane, and Cope 1957). Since that time, it has been realized that hypercalcemia from any cause, including metastatic osteolysis and multiple myeloma, can be associated with symptomatic hypomagnesemia, because hypercalcemia causes increased excretion of magnesium in the urine (Eliel et al. 1968).

That the two major intracellular cations, potassium and magnesium, behave similarly was originally noted in the course of diabetic ketoacidosis (Martin and Wertman 1947). Similarities between the two ions were further emphasized by the observation that serum potassium and magnesium are both low in cases of primary aldosteronism (Mader and Iseri 1955). Moreover, cardiac arrhythmias, such as ventricular extrasystoles, ventricular tachycardia, and ventricular fibrillation, similar to those induced by hypokalemia, have been described in various clinical disorders known to result in depletion of magnesium. As a matter of fact, the demonstration that in hypertensive patients treated with diuretics, potassium supplementation alone has little effect on ventricular extrasystoles unless magnesium supplementation is added to the regimen has aroused clinical interest in magnesium deficiency significantly over recent years (Hollifield 1984).

Controversial Issues

A Puzzling Syndrome

During the 1960s, human magnesium deficiency was established as a clinical entity beyond any doubt. However, conflicting reports in the literature may

have created the view for many that clinical magnesium deficiency is a puzzling syndrome comprising almost every symptom and sign. Such confusion also results partly from the previously mentioned difficulties in assessing magnesium status along with superimposed nonspecific manifestations related to primary illness or other concomitant metabolic abnormalities.

Thus, in addition to the generally accepted signs and symptoms of neuromuscular hyperactivity and disturbance of cardiac rhythm (Table IV.B.4.4), a great variety of manifestations, ranging from difficulty in learning to multiple phlebothromboses, have been proposed as components of the clinical picture of magnesium deficiency (Table IV.B.4.5). Moreover, the observation that even severe magnesium deficiency may have no symptoms or signs has added significantly to the confusion (Martin, Mehl, and Wertman 1952).

The term "spasmophilia" was first introduced by the French physician Nicolas Corvisart in 1852. In 1874, Wilhelm Erb attributed spasmophilia to neuromuscular hyperexcitability, and since then the term has been used to denote "a condition in which the motor nerves show abnormal sensitivity to mechanical or electric stimulations and the patient shows a tendency to spasm, tetany and convulsions" (Dorland 1943).

In 1959, Jean Durlach described a syndrome that he called "hypomagnesemic constitutional spasmophilia." The syndrome, which was considered to be "the nervous form of primary chronic magnesium deficiency in adults," comprised a long list of nonspe-

Table IV.B.4.5. *Additional symptoms and signs attributed to magnesium deficiency by some authors*

Neuromuscular and psychologic manifestations

Anxiety
Depression, hysteria, hypochondria (neurotic triad)
Intellectual and affective retardation
Learning deficiencies
Headaches
Dizziness
Insomnia
Sight weakness
Startle response and noise sensitivity
Neck and back pains
Acroparesthesia
Raynaud's syndrome
"Spasmophilia"

Cardiac manifestations

Coronary artery spasm
Greater size of myocardial infarction

Miscellaneous

Hair and ungual brittleness
Dysplasia of teeth enamel
Opacification of lens
Ostomalacia
Gastritis, visceral spasm, or dyskinesia ("visceral spasmophilia")
Mitral valve prolapse

Source: Data from Durlach (1988a).

Table IV.B.4.4. *Generally accepted symptoms and signs of magnesium deficiency*

Neuromuscular manifestations

Muscular twitching and tremor
Myoclonic jerks
Convulsions
Ataxia and nystagmus
Irritability and restlessness
Dysphagia
Dysarthria
Chvostek's sign (and rarely Trousseau's sign)
Tetany (rarely)
Apathy
Coma

Cardiac manifestations

Premature ventricular contractions
Atrial fibrillation
Torsades de pointes
Ventricular fibrillation
Sudden death

Metabolic effects

Refractory hypokalemia
Hypocalcemia

Source: Data from Wacker and Parisi (1968) and Flink (1980).

cific neuromuscular, psychological, cardiovascular, and gastrointestinal manifestations, such as anxiety, excessive emotivity, globus hystericus, dyspnea *sine materia*, dizziness, insomnia, headaches, myalgias, cramps, tremors, tetanic attacks, palpitations, Raynaud's syndrome, biliary dyskinesia, epigastric cramps, and syncope. Symptoms were ameliorated by oral administration of magnesium salts. Low erythrocyte magnesium, a positive Chvostek's sign, and abnormal electromyographic, electroencephalographic, and electronystagmographic tracings were reported as characteristic of the syndrome (Durlach and Lebrun 1959).

In the years that followed, the compass of clinical magnesium deficiency was broadened to encompass such divergent manifestations as phlebothrombosis, allergic disorders, intellectual retardation, sight weakness, hepatic dysfunction, cases of dysmenorrhea, spontaneous abortion, and mitral valve prolapse. This broadening of the spectrum of clinical manifestations of magnesium deficiency was based on a great number of cases studied by Durlach and other investigators in continental Europe.

The fact that publication of such cases in English has been rare is indicative of the reluctance of American and British experts to accept the existence of a syndrome such as "idiopathic spasmophilic diathesis" due to magnesium deficiency. However, an incident investigated by the U.S. Food and Drug Administration (FDA) in 1980 and a recent double-blind trial performed in Southampton imply that the idea of the

involvement of magnesium deficiency in bizarre clinical syndromes should not be totally rejected.

The incident investigated by the FDA was as follows: 11 high school football players with diet histories suggesting low calcium and magnesium intake were given a phosphate-free, soluble calcium preparation because of leg cramps. They were then subjected to the noise and physical contact of a football game. Over a short period, 8 of them developed serious neuromuscular dysfunction manifested by ataxia, slurred speech, hyperventilation, muscle spasm, and tonic-clonic seizures. All of them recovered without specific treatment, and none developed neurologic sequelae. About 10 years later, W. F. Langley and D. Mann noticed the similarity of this syndrome with grass tetany in animals and postulated that central nervous system magnesium deficiency was responsible for its occurrence. The condition was thought to be brought on by a sudden rise in serum calcium and was termed "reactive symptomatic magnesium deficiency" (Langley and Mann 1991).

The randomized double-blind, placebo-controlled trial in Southampton concerned the effect of intramuscular administration of magnesium sulfate on the symptoms of chronic fatigue syndrome (Cox, Campbell, and Dowson 1991). As stated by the authors, the study was undertaken because many of the symptoms of the chronic fatigue syndrome – anorexia, nausea, learning disability, personality changes, tiredness, and myalgia – are similar to those of magnesium deficiency, and because in a pilot study they found that patients with the syndrome had low red blood cell magnesium concentrations. Within six weeks, red cell magnesium had returned to normal in all patients treated with magnesium. In addition, magnesium administration resulted in improvement in energy, pain perception, emotional reactions, sleep patterns, sense of social isolation, and physical mobility, as scored with the use of the Nottingham health profile. The authors confessed, however, that their trial was small and that there was only 6 weeks of follow-up and, therefore, the results were hardly conclusive (Cox, Campbell, and Dawson 1991).

The "Water Story"

In 1957, J. Kobayashi reported a geographical relationship between stroke-associated mortality and river water acidity in Japan (Kobayashi 1957). This was the beginning of what was later called the "water story." Following the publication of the Japanese data, H. A. Schroeder (1960) investigated the relationship between mortality from different diseases and mineral content of drinking water in the United States, and concluded that the important determinant was water hardness rather than water acidity. Drinking soft water, he suggested, may promote the prevalence of cardiovascular disease, whereas hard water may exert a protective action. Since then, associations between hardness of drinking water and mortality from cardiovascular dis-

ease have been reported in England and Wales, in Sweden, and in Ontario – the softer the water the higher the mortality from cardiovascular disease.

Ten years after the original report of Kobayashi, T. Crawford and M. D. Crawford (1967) compared cardiac lesions found in medicolegal necropsies in cases of sudden death in two areas: Glasgow, a notable soft-water area with a high cardiovascular disease mortality; and London, a city with very hard water and a considerably lower mortality from cardiovascular causes. They realized that despite the large difference in mortality, the incidence of coronary atherosclerosis was similar in both places. This finding was subsequently interpreted to mean that the excess mortality in the soft-water area might correspond to sudden deaths due to fatal ventricular arrhythmias. Moreover, chemical analysis of the coronary arteries revealed very low values for calcium and magnesium in the soft-water area, suggesting that the mineral content of the arteries was related to the mineral content of the drinking water.

On the basis of the evidence just mentioned, a "cardiac water factor" has been postulated to exist. Presumably, such a factor would be either something beneficial in hard water or something harmful in soft water. Since the bulk of water hardness is made up of calcium and magnesium, these two ions have been the more likely candidates for the "cardiac water factor." Magnesium is a much more abundant intracellular ion than calcium. Furthermore, a considerably greater proportion of the daily magnesium intake, as compared with the daily calcium intake, comes from drinking water. Coupled with the arrhythmogenic potential of magnesium depletion, these facts have led to the hypothesis that magnesium is the cardio-protective factor contained in hard water.

In support of this hypothesis, T. W. Anderson and colleagues (1975) found that magnesium concentrations in myocardial samples obtained from accident victims were significantly lower in five soft-water cities of Ontario than in three hard-water cities of the same province. However, a similar report in England at about the same time (Chipperfield et al. 1976) found a significant difference in mean myocardial magnesium concentration between heart muscle samples obtained from noncardiac deaths in two cities with different water hardness, but the difference was "in the wrong direction." This discrepancy led Anderson and colleagues to conclude that if there was any sort of 'water factor' in Britain, it probably involved something beside magnesium (Anderson et al. 1980).

In the meantime, other simple correlation or multivariate statistical studies, including analysis of the data of the World Health Organization myocardial infarction registry network, have failed to discover a causal relationship between water hardness and heart disease mortality. Schroeder's report has been criticized as misleadingly simplistic, because it was based on correlations between just one index of environmental

exposure and a set of death rates (Hammer and Heyden 1980). In addition, the demonstration that only a small proportion of the daily mineral element intake comes from drinking water has cast doubt on the possible physiological significance of water hardness.

Advocates of the "water factor," however, have adopted a somewhat different approach that emphasizes not water hardness but the interrelationship of individual elements contained in drinking water, along with the intake of these elements as they are ingested with food (Marier 1981). In this context, of particular interest are the observations of H. Karppanen in Finland, a country with a very high prevalence of cardiovascular disease. Mortality from ischemic heart disease in Finland shows a peculiar geographic distribution: Starting from the eastern areas, where it is extremely high, it decreases continuously toward the western and southwestern part of the country. These regional differences in mortality correlate with differences in the content of exchangeable magnesium in the arable soils: Magnesium concentration in the soil of eastern areas is only about one-third of that of the southwestern areas. Soil is the primary source of magnesium. Low magnesium concentration in soil is, therefore, expected to result in a low magnesium concentration in drinking water, and also in a low magnesium content of the cereal crops, which are a major food source of magnesium.

In addition to these observations, Karppanen has reported a strong positive correlation between the death rate from ischemic heart disease and the estimated average calcium-to-magnesium ratio of the diet in various Organization for Economic Cooperation and Development (OECD) countries. According to Karppanen, this correlation suggests that what really matters in the "water story" is the ratio of calcium-to-magnesium intake rather than the sum or the absolute amount of each of these elements (Karppanen 1981).

Nutrient or Drug?

Magnesium was known as a drug long before it was recognized as a nutrient. Thus, even in the nineteenth century, parenteral magnesium sulfate was used in the management of eclamptic convulsions and also as an ancillary anesthetic agent. In cardiology, magnesium was first introduced in 1935, when L. Zwillinger reported a beneficial effect of intravenous magnesium sulfate on paroxysmal tachycardia and extrasystolic arrhythmia in patients treated with digitalis. Since then, parenteral magnesium salts have been repeatedly used as an antiarrhythmic agent. Intravenous magnesium has also been recognized as a potent vasodilator agent, useful in the treatment of heart failure and angina pectoris.

During the past five years, several randomized clinical trials have examined the effects of intravenous magnesium in acute myocardial infarction. Collectively, these trials strongly suggest that magnesium

therapy may result in decreased mortality as well as the frequency of postinfarction arrhythmias. A larger trial with a planned sample size of 2,500 patients was started in 1987 (Woods 1991).

In all these studies and in many of the cases of treatment of cardiac arrhythmias, the beneficial effect of intravenous magnesium has been related to pharmacological action rather than to repletion of a deficit. In fact, when magnesium is given intravenously, serum magnesium levels are usually raised well above normal values. However, hypomagnesemia is an established cause of cardiac arrhythmias, and magnesium deficiency has been shown to be associated with sudden cardiac death and increased mortality in the acute stage of myocardial infarction (Dyckner 1980). Moreover, low levels of magnesium have been found in myocardial tissue obtained at necropsy from patients with acute myocardial infarction. Consequently, under conditions of magnesium deficiency, a supplement of magnesium is expected to have the same favorable effect on cardiac arrhythmias, and possibly the same protective action against myocardial ischemia, as the administration of magnesium salts in pharmacological doses.

On the basis of the evidence just mentioned, it is clear that the characterization of magnesium as a drug or as a nutrient depends on the distinction between pharmacological action and replenishment of deficient stores (Charbon 1989). When there is a proven magnesium deficiency, magnesium acts as a nutrient. The patient is then expected to be completely cured as soon as the deficit is corrected. However, parenteral magnesium salts may be given as a drug. In this case, though, their potency and duration of action will depend on magnesium concentration in plasma, and their effect will be the same in each case of repeated administration.

Adequacy of Intake

Estimates of the daily requirements of magnesium have been largely based on the original calculations published in 1942 by J. Duckworth and G. Warnock, who calculated the magnesium requirements by analyzing the available data from balance determinations done in the 1930s. Currently, the Recommended Dietary Allowance (RDA) proposed by the Food and Nutrition Board of the National Research Institute of the United States is 350 mg per day for men and 300 mg per day for women, providing about 4.5 to 5.0 milligrams (mg) per kilogram of body weight per day (Munro 1980). RDA is increased to 450 mg per day during pregnancy and lactation and to 400 mg per day during adolescence in males. For infants and children the corresponding values are 100 and 200 mg per day.

Magnesium is so plentiful in both plant and animal foods that the recommended daily intake is readily obtained in ordinary diet. Common dietary sources include unprocessed cereal, nuts, legumes, vegetables, seafood, dairy products, meats, and drinking water.

However, Mildred Seelig (1964) in the United States and Durlach (1988a) in France have suggested that magnesium intake in developed countries may be insufficient to meet daily needs. Since then this concern has been shared by many nutritionists.

The shortage of magnesium in contemporary diets has been attributed to the increased reliance of people in most Western countries on highly purified foods, such as sugar, starch, soft drinks, and distilled alcohol, that contain very little magnesium. Agricultural techniques of accelerated growth, resulting in decreased magnesium fixation by plants, and the use of magnesium-poor soil fertilizers, and also of pesticides that inhibit magnesium absorption, are considered as additional causes of decreased magnesium content in food items currently available for civilian consumption in industrialized countries. In fact, analysis of sample meals in the United States, Canada, and several European countries has revealed that the contemporary magnesium intake ranges from minimal adequacy to as low as 50 percent of the RDA (Table IV.B.4.6). This observation seems to hold particularly true for special population groups, such as pregnant women, teenage girls, or elderly people, because of increased needs, limited nutritional intake, or the age-associated decline in intestinal magnesium absorption (Mountokalakis 1987). In view of these considerations, Seelig has suggested that the recommendations for daily magnesium intake should be revised upward to as much as 7 to 10 mg/kg body weight.

Other experts opposed to this view have argued that since the RDA includes a generous margin of safety above the current estimates of the minimal daily requirements, the expression of adequacy of magnesium intake as a proportion of the RDA may be misleading. As a matter of fact, metabolic studies have indicated that a positive magnesium balance can be maintained at intakes as low as 60 percent of the currently proposed RDA. The answer to these arguments has been that because RDA is calculated by analyzing the results of metabolic studies conducted under conditions of relative serenity, it does not take into account possible changes of magnesium requirements related to the stresses of everyday life. The advocates of the "magnesium malnutrition" hypothesis emphasize that although the reported suboptimal intakes do not necessarily lead to overt clinical magnesium deficiency, they may represent a "long-term marginal magnesium insufficiency" with the potential risk of increased vulnerability to several disease processes (Marier 1982). The present uncertainty about evaluating the clinical status of magnesium impedes any attempt to clarify this issue further.

Epilogue

The story of magnesium is a story of discovery and also a story of neglect. Over the past few decades, enthusiasm about magnesium has repeatedly turned

Table IV.B.4.6. *Magnesium intake in the modern-day world*

Locality and special group (range)	Year	Population studied	Intake as percentage of RDA (average of)
Healthy adults			
United States			
Central states	1961	30 women	54.0–93.0
Boston	1970	955 men	74.9
Canada			
Montreal	1978	15 men	70.6
Montreal	1978	15 women	74.7
Newfoundland	1978	83 men	53.4
Newfoundland	1978	105 women	47.7
United Kingdom	1972	–	76.9
France	1980	–	80.0–100.0
West Germany	1972	1,852 adults	72.3
Special population groups			
Institutionalized people			
United States	1969	men	62.7
Indiana	1977	31 aged men	54.3–71.7
Indiana	1977	34 aged women	54.0–94.3
Belfast	1979	36 aged men	56.6
Belfast	1979	90 aged women	56.3
Pregnant women			
United States	1978	–	45.3
United States	1979	aged 19–29	60.0
Girls			
Indiana	1976	80, aged 12–14	79.3
Indiana	1977	76, aged 12–14	77.0
Tennessee	1980	60, aged 9–11	66.0–100.0

Source: Data from Marier (1982).

to indifference. Since 1971, when the First International Symposium on Magnesium Deficit in Human Pathology was held at Vittel (France), five international congresses and several American and European meetings on magnesium have taken place, and three journals exclusively devoted to magnesium have been published. Yet despite all this worldwide research activity, the practicing physician has relatively little interest in magnesium, and hypomagnesemia is currently the most underdiagnosed electrolyte deficiency (Whang 1987). Moreover, essential data for diagnosis and physiopathology are still lacking.

No doubt magnesium is an important intracellular element with a key role in many metabolic functions. The irony is that this multifaceted role of magnesium in the organism has been the main obstacle in understanding its fundamental function. A characteristic example is the physiological interrelationships of magnesium and hormones. Companion to both calcium and potassium, magnesium has long been assumed to interact with the hormones involved in the homeostasis of these two ions. The evidence, however, now shows that neither parathyroid hormone nor aldosterone

terone exerts an overriding control on magnesium metabolism. Yet the facts that magnesium and calcium are mutually influenced and that hypomagnesemia can cause loss of intracellular potassium, in spite of a normal plasma potassium, are not in question.

Some 30 years ago, magnesium was hardly mentioned in the medical textbooks. Since then, much more information has been acquired. Although most of the current textbooks contain adequate data on the subject, they repeatedly state that magnesium is found in so many foods that, ordinarily, magnesium deficiency is rare in healthy people. It should be noted, however, that although magnesium is well conserved by both the kidneys and the bowel when the supply is limited, it is poorly stored so that a regular intake is needed to avoid deficiency. It is, therefore, natural to assume that a marginal deficiency state may be not uncommon. And once it was realized that magnesium has important cardiovascular effects, emphasis shifted from a focus on overt clinical magnesium deficiency to interest in the concept of marginal magnesium deficiency as a risk factor for cardiovascular disease. The importance of this concept is underscored by the characterization of magnesium as "nature's physiological calcium blocker" (Iseri and French 1984).

Because most of the body's magnesium is intracellular, the challenge for the future is to develop a feasible test in clinical medicine that will give meaningful information on the overall intracellular magnesium status. The question of intracellular magnesium is, without doubt, the most difficult to resolve. It is, however, the clarification of this issue that will enable us to understand better the relationship of magnesium to health and disease and to ascertain the magnitude of the segment of the population that may be deficient in this important nutrient.

Theodore D. Mountokalakis

Bibliography

- Aikawa, J. K. 1981. *Magnesium: Its biologic significance*. Boca Raton, Fla.
- Alcock, N. W. 1969. Development of methods for the determination of magnesium. *Annals of the New York Academy of Sciences* 162: 707-16.
- Anderson, T. W., W. H. Leriche, D. Hewitt, and L. C. Neri. 1980. Magnesium, water hardness, and heart disease. In *Magnesium in health and disease*, ed. M. Cantin and M. S. Seelig, 565-71. New York.
- Anderson, T. W., L. C. Neri, G. B. Screiber, et al. 1975. Ischemic heart disease, water hardness and myocardial magnesium. *Canadian Medical Association Journal* 113: 199-203.
- Barnes, B. A., S. M. Krane, and O. Cope. 1957. Magnesium studies in relation to hyperparathyroidism. *Journal of Clinical Endocrinology and Metabolism* 17: 1407-21.
- Charbon, G. A. 1989. Magnesium treatment of arrhythmia: Drug or nutritional replenishment? Pitfalls for the experimental design. In *Magnesium in health and disease*, ed. Y. Itokawa and J. Durlach, 223-8. London.
- Chipperfield, B., J. R. Chipperfield, G. Behr, and P. Burton. 1976. Magnesium and potassium content of normal heart muscle in areas of hard and soft water. *Lancet* 1: 121-3.
- Cox, I. M., M. J. Campbell, and D. Dowson. 1991. Red blood cell magnesium and chronic fatigue syndrome. *Lancet* 337: 757-60.
- Crawford, T., and M. D. Crawford. 1967. Prevalence and pathological changes of ischaemic heart disease in a hard-water and in a soft-water area. *Lancet* 1: 229-32.
- Dorland, D. A. N. 1943. *The American illustrated medical dictionary*. Ninth edition. Philadelphia, Pa.
- Duckworth, J., and G. Warnock. 1942. The magnesium requirements of man in relation to calcium requirements with observations on the adequacy of diet in common use. *Nutritional Abstracts and Reviews* 12: 167-83.
- Durlach, J. 1988a. *Magnesium in clinical practice*. London.
- 1988b. Magnesium research: A brief historical account. *Magnesium Research* 1: 91-6.
- Durlach, J., and R. Lebrun. 1959. Magnésium et pathogénie de la spasmophilie constitutionnelle idiopathique. *Comptes Rendues de la Société de Biologie* 153: 1973-5.
- Dyckner, T. 1980. Serum magnesium in acute myocardial infarction. *Acta Medica Scandinavica* 207: 59-66.
- Eliel, L. P., W. O. Smith, R. Chanes, and J. Howrylko. 1969. Magnesium metabolism in hyperparathyroidism and osteolytic disease. *Annals of the New York Academy of Sciences* 162: 810-30.
- Flink, E. B. 1980. Clinical manifestations of acute magnesium deficiency in man. In *Magnesium in health and disease*, ed. M. Cantin and M. S. Seelig, 865-82. New York.
1985. Magnesium deficiency in human subjects - a personal historical perspective. *Journal of the American College of Nutrition* 4: 17-31.
- Flink, E. B., et al. 1954. Magnesium deficiency after prolonged parenteral fluid administration and after chronic alcoholism complicated by delirium tremens. *Journal of Laboratory and Clinical Medicine* 43: 169-83.
- Forster, J. 1873. Beiträge zur Ernährungsfrage. *Zeitschrift für Biologie*.
- Fraser, R., and E. B. Flink. 1951. Magnesium, potassium, phosphorus, chloride and vitamin deficiency as a result of prolonged use of parenteral fluids. *Journal of Laboratory and Clinical Medicine* 38: 817-29.
- Gaube, J. 1895. *Théorie minérale de l'évolution et de la nutrition de la cellule animale*. Paris.
- Hammer, D. I., and S. Heyden. 1980. Water hardness and cardiovascular mortality. An idea that has served its purpose. *Journal of the American Medical Association* 243: 2399-400.
- Harman, M. 1956. Parathyroid adenoma in a child. *Journal of Diseases of Children* 91: 313-56.
- Hippocrates. 1988. Internal affections. In *Hippocrates*, trans. P. Potter, Vol. 6, 140-1. Cambridge, Mass.
- Hirschfelder, A. B., and V. G. Hauray. 1934. Clinical manifestations of high and low plasma magnesium. Dangers of Epsom salt purgation in nephritis. *Journal of the American Medical Association* 102: 1138-41.
- Hollifield, J. W. 1984. Potassium and magnesium abnormalities: Diuretics and arrhythmias in hypertension. *The American Journal of Medicine* 77: 28-32.
- Iseri, L. T., and J. H. French. 1984. Magnesium: Nature's physiological calcium blocker. *American Heart Journal* 108: 188-93.
- Jolyet, F., and M. Cahours. 1869. Sur l'action physiologique des sulfates de potasse, de soude et de magnésie en

- injection dans le sang. *Archives de Physiologie Normale et Pathologique* 2: 113–20.
- Karppanen, H. 1981. Epidemiological studies on the relationship between magnesium intake and cardiovascular diseases. *Artery* 9: 190–9.
- Kobayashi, J. 1957. Geographical relationship between the chemical nature of river water and death-rate from apoplexy. *Ber Obara Instituts Landwirtschaftliche Biologie* 11: 12–21.
- Kruse, H., E. R. Orent, and E. V. McCollum. 1932. Magnesium deficiency in animals. *Journal of Biological Chemistry* 96: 519–32.
- Kruse, H., M. Schmidt, and E. McCollum. 1934. Studies on magnesium deficiency in animals. *Journal of Biological Chemistry* 106: 553–72.
- Kubota, J. 1981. How soils and climate affect grass tetany. *Crops and Soils Magazine* 33: 15–17.
- Langley, W. F., and D. Mann. 1991. Central nervous system magnesium deficiency. *Archives of Internal Medicine* 151: 593–6.
- Leroy, J. 1926. Nécessité du magnésium pour la croissance de la souris. *Comptes Rendus de la Société de Biologie* 94: 431–3.
- Loeb, B. 1900. On the different effect of ions upon myogenic and neurogenic rhythmical contractions and upon embryonic and muscular tissue. *American Journal of Physiology* 3: 383–96.
- Mader, I. J., and L. T. Iseri. 1955. Spontaneous hypopotassemia, hypomagnesemia, alkalosis and tetany due to hypersecretion of corticosterone-like mineralocorticoid. *The American Journal of Medicine* 19: 976–88.
- Marier, J. R. 1981. Water hardness and heart disease. *Journal of the American Medical Association* 245: 1316–26.
1982. Quantitative factors regarding magnesium status in the modern-day world. *Magnesium* 1: 3–15.
- Martin, H. E., J. Mehl, and M. Wertman. 1952. Clinical studies of magnesium metabolism. *Medical Clinics of North America* 36: 1157–71.
- Martin, H. E., and M. Wertman. 1947. Serum potassium, magnesium and calcium levels in diabetic acidosis. *Journal of Clinical Investigation* 26: 217–28.
- Mendel, L. B., and S. R. Benedict. 1909. The paths of excretion for inorganic compounds. IV. The excretion of magnesium. *American Journal of Physiology* 25: 1–7.
- Miller, J. F. 1944. Tetany due to deficiency in magnesium. Its occurrence in a child of six years with associated osteochondrosis of capital epiphysis of femurs. *American Journal of Diseases of Children* 67: 117–19.
- Mountokalakis, T. 1987. Effects of aging, chronic disease, and multiple supplements on magnesium requirements. *Magnesium* 6: 5–11.
- Munro, H. N. 1980. *Recommended dietary allowances*. Ninth edition, 114–36. Washington, D.C.
- Schroeder, H. A. 1960. Relations between mortality from cardiovascular disease and treated water supplies. *Journal of the American Medical Association* 117: 1902–8.
- Seelig, M. S. 1964. The requirements of magnesium by normal adults. *American Journal of Clinical Nutrition* 14: 342–90.
- Sjollema, B. 1932. Nutritional and metabolic disorders in cattle. *Nutrition Abstracts and Reviews* 1: 621–32.
- Sjollema, B., and L. Seekles. 1932. Der Magnesiumgehalt des Blutes. Besonders bei Tetanie. *Klinische Wochenschrift* 11: 989–90.
- Smith, R. H. 1959. Absorption of magnesium in large intestine of calf. *Nature* 184 (Suppl. 11): 821–6.
- Wacker, W. E. C. 1969. The biochemistry of magnesium. *Annals of the New York Academy of Sciences* 162: 717–26.
- Wacker, W. E. C., and A. F. Parisi. 1968. Magnesium metabolism. *The New England Journal of Medicine* 278: 712–17.
- Whang, R. 1987. Magnesium deficiency: Pathogenesis, prevalence, and clinical implications. *The American Journal of Medicine* 82 (Suppl. 3A): 24–9.
- Willstätter, R., and A. Stoll. 1913. *Untersuchungen über Chlorophyll*. Berlin.
- Woods, K. L. 1991. Possible pharmacological actions of magnesium in acute myocardial infarction. *British Journal of Pharmacology* 32: 3–10.
- Zwillinger, L. 1935. Ueber die Magnesiumwirkung auf das Herz. *Klinische Wochenschrift* 14: 1419–33.

IV.B.5 🐉 Phosphorus

Phosphorus (abbreviated as P) is a highly toxic element, which, when it occurs in the form of a phosphate salt, becomes a nutrient essential for human health. Phosphates are found in practically every type of food and, consequently, are plentiful in the typical diet. Inorganic phosphates (abbreviated as P_i) are absorbed from food as electrically charged salt anions. The organic phosphates (P_o) that exist in cells and extracellular compartments of foods are primarily converted to P_i through digestive processes prior to absorption. A few organic phosphates are apparently absorbed as part of small fat-soluble organic molecules, such as the phosphates of phospholipids. The concentration of P_o molecules, however, is not under homeostatic control, in contrast to the concentration of P_i , which is regulated along with calcium (Ca) in blood and extracellular fluids.

The close association between calcium and P_i in the extracellular body fluids and in bone tissue requires joint consideration of dietary calcium and dietary phosphates for an understanding of their physiological linkages and the important relationship between low calcium intakes and high phosphate intakes. This relationship potentially contributes to altered calcium homeostasis and the loss of bone mass.

Several aspects of P_i are reviewed here in an attempt to place these essential chemical anions in the perspective of their utilization in human health and disease. The physiological functions of P are reviewed first, and a number of general references have also been included in the bibliography for the interested reader.

Physiological Functions of Phosphates

The regulation of P_i in the blood is maintained by a complex homeostatic system that also controls the blood concentration of Ca. P_i ions can follow several potential metabolic pathways after entry into cells.

One immediate use of P_i ions is to phosphorylate glucose through enzymatic steps via kinases. They are also used in the phosphorylation of several other molecules; for example, creatine phosphate serves as an energy reserve in skeletal muscle tissue, and adenosine triphosphate is the primary donor of energy from its high-energy bonds within cells. In addition, several types of phospholipids and nucleic acids incorporate P_i within their molecular structures.

Cellular uses of P in intermediary metabolism are very extensive. Practically all energetic steps utilize high-energy phosphate bonds (adenosine triphosphate, or ATP) for the synthesis of organic molecules: to drive transport systems across cell membranes, to make muscles contract, to allow nerves to conduct impulses and transfer information, to convey genetic information, and to provide skeletal support and protection. In addition, phosphates circulating in blood have buffering activity. Clearly, P_i ions have multiple uses both within and without cells. Mineralized bone serves as an important store of P_i ions that can be retrieved through the action of parathyroid hormone (PTH) on bone cells. Because of the large reservoir of P_i ions, hypophosphatemia and P deficiency are rare events in adults without other major complications.

Phosphates in Foods

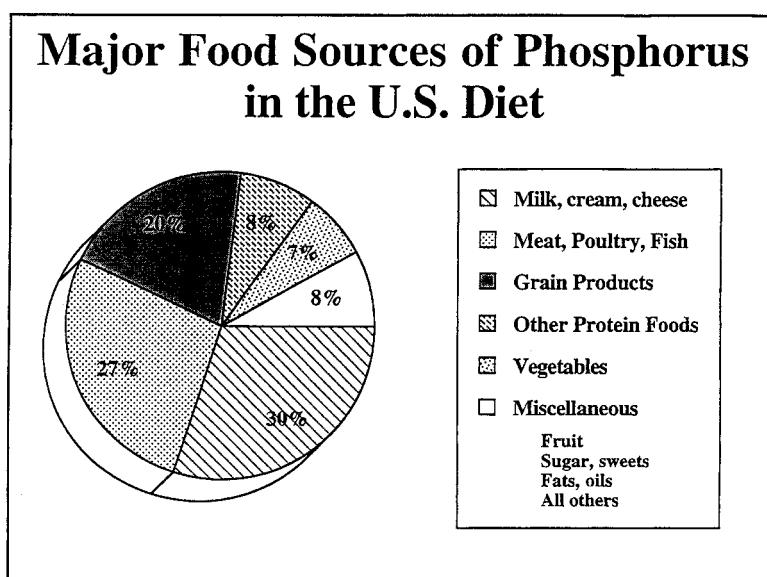
In terms of dietary sufficiency, P does not present a problem for human health because of the abundance of molecules containing this element in the food supply. Rather, the potential health problem is dietary Ca insufficiency in relation to excess intake of phosphates over periods of years, or even decades. Phosphorus deficiency, a rare clinical disorder, occurs almost exclusively because of a pathological change

in the handling of phosphates, rather than because of a dietary inadequacy. The reason for these encompassing statements is that almost all foods contain phosphate groups in both organic (P_o) and inorganic (P_i) forms, and many of the foods – especially animal products – commonly consumed by all populations of the world are rich in P.

The approximate percentage distribution of foods (by food group) that provide dietary P in the United States is as follows: Milk and dairy products – 30 percent; meat, poultry, and fish – 27 percent; cereal grains and grain products – 20 percent; legumes, nuts, and seeds – 8 percent; vegetables – 7 percent; and other (miscellaneous) foods, including fruits – 8 percent (Figure IV.B.5.1). Although accurate data are lacking, it is presumed that these percentages are similar for the nations of the European Union, along with other Western countries. For populations that consume few dairy foods (or none), the percentage of dietary P contributed by grains, legumes, vegetables, and fruits would be greatly increased, depending on the food traditionally available. Vegetarians of all types, but especially “vegans” (strict vegetarians), have lower P intakes than do omnivores who consume several servings a day of dairy foods and meats. But, even though P intakes by vegans could be insufficient, deficiency symptoms are very unlikely.

As Figure IV.B.5.1 illustrates, P is widely found in foods in their natural unprocessed condition, and available food composition tables report the amounts of P measured in natural foods without any phosphate additives. Unfortunately, no food tables exist for the P content of processed foods. This lack is a serious hindrance to the accurate estimation of total phosphorus consumption from all foods, processed and unprocessed.

Figure IV.B.5.1. Approximate percentage contributions of the major food groups to the consumption of phosphorus. (Note: Phosphate additives are not included in these estimates.) (From USDA, CSFII 1994; adapted from Anderson and Garner 1996.)



Plant Foods

Although cereal grains and most vegetables yield good amounts of phosphates, these are nonetheless typically smaller amounts than those provided by animal foods. Moreover, except for highly processed wheat flours and polished rice, cereal grains have much of their P_o bound in phytates that are not completely digested within the gastrointestinal (GI) tract, and therefore, the total amount of P in grains is not available for absorption. Legumes and foods made from them, including soy flour and peanuts, are good sources of P, but this is not the case for most other plant foods, save for nuts such as almonds. Fruits, fruit juices, and vegetable oils contain negligible amounts of P. Table IV.B.5.1 lists the amounts of P (as a combination of both P_i and P_o) and calcium in commonly consumed foods.

Cereal grains contain significant amounts of the polyphosphate phytic acid (inositol hexaphosphate) in the bran and other parts. Each phytate molecule contains 6 P_o groups that become P_i groups when phytates are digested within the lumen of the GI tract by phosphatase enzymes. Food composition tables should be consulted for the phosphorus (combined P_o and P_i) content of specific foods.

Table IV.B.5.1. *Content of phosphorus and calcium in commonly consumed food in mg per serving*

Food Item	Serving size (1)	Phosphorus, mg	Calcium, mg	Calcium: phosphorus ratio
Yoghurt, plain	cup	326	415	1:0.8
Milk, 3.5%	cup	228	291	1:0.8
Milk, skim	cup	247	302	1:0.8
Cheese, Amer.	1 oz	211	174	1:1.2
Mozzarella, skim	1 oz	149	207	1:0.7
Egg, whole	large	86	25	1:3.4
Beef, ground	3 oz	135	8	1:16.9
Pork chop, loin	one	184	8	1:23
Chicken, drumstk	one	81	13	1:6.2
Turkey, white mt	3 oz	186	16	1:11.6
Tuna, water pack	3 oz	103	10	1:10.3
Corn flour	cup	318	8	1:40
Wheat flour	cup	135	19	1:7.1
Oatmeal	0.5 cup	89	9	1:9.9
Bread, whole wheat	slice	74	20	1:3.7
Bread, white	slice	30	35	1:0.9
Potato, white	one	115	21	1:5.5
Potato, sweet	one	63	32	1:2.0
Peas, green	0.5 cup	72	19	1:3.8
Baked beans	0.5 cup	132	64	1:2.1
Broccoli	0.5 cup	50	47	1:1.1
Corn, kernel	0.5 cup	39	2	1:19.5

Source: Hands, FOOD Finder: Food Sources of Vitamins and Minerals (1990).

Animal Foods

Animal foods are rich in P, with eggs, meats, fish, crustaceans, mollusks, poultry, and dairy products containing large amounts of this element. Liver, cheeses, and eggs (yolks) are highest in phosphorus, followed by meats, fish, and poultry, and then by milks (of any fat content). (See Table IV.B.5.1 for a listing of the P content of commonly consumed foods.) Mixed dishes that feature cheese or milk also have large amounts of P.

Food Additives

Phosphates (both P_i and P_o) are increasingly found in the food supplies of economically developed nations because of their widespread use as chemical additives in food applications. The food and beverage industries add a significant amount of phosphates to many foods and to cola-type beverages. Phosphate additives represent a diverse group of molecules, practically all of which are readily solubilized within the stomach or upper small intestine and are thereby highly bioavailable. Processed foods currently consumed by North Americans often have a significantly increased phosphorus content because of the intentional addition of one or more of these phosphate salts (Calvo 1993; Calvo and Park 1996). Phosphate additives have several functions in foods, notably as acids, buffers, anti-caking agents, emulsifiers, and sequestrants (Dziezak 1990), and individuals who consume many processed foods (especially those with cheese in them) are estimated to have an increased daily phosphorus intake of 10 to 15 percent.

Supplements

Because practically all foods naturally contain phosphorus, and many processed foods have phosphates added to them, no need for supplementation exists in healthy individuals with normal renal function. Except for very rare clinical cases of renal phosphate wasting condition, or phosphate deficiency in newborns or premature babies, it is difficult to conceive of a situation requiring phosphate supplements.

Phosphate Additives – Recent Changes

The use of phosphate additives in food processing has accelerated since 1950, and in fact, for the first time in history, humankind faces an excess of phosphorus in the food supply because of the intentional addition of an estimated 40 to 50 different phosphate salts (Calvo and Park 1996). This, in turn, has introduced the possibility of an adverse effect of total phosphate consumption (from foods, beverages, and additives) on bone health, much as excess sodium intake from sodium-processed foods has contributed to the high prevalence of hypertension in the United States and other nations.

An interesting aspect of this increased intake of P is that it is not included in national survey esti-

mates of total P consumption, meaning that the true intake of phosphorus in the United States is substantially underestimated (Oenning, Vogel, and Calvo 1988; Calvo and Park 1996). A study by M. S. Calvo and Y. K. Park (1996) shows a trend of increasing availability of phosphate additives in the food supply from specific items, such as frozen pizza, frozen processed poultry, and frozen prepared foods. Particularly worrisome is the increased availability of frozen processed poultry products in the marketplace over the last few years. (Frozen-food manufacturers, especially, utilize phosphate additives because of the phosphates' stability when the foods are thawed or heated.) Fast-food entrees also often contain phosphate additives, but little or none of the phosphate-additive content of these foods is included in estimates of phosphorus consumption (Calvo and Park 1996).

Dietary Intakes of Phosphates

Current dietary patterns in the United States suggest that, on the average, American women are consuming too little Ca in relation to P. In every age category over 11 years, the calcium intakes of females are so low that the median Ca:P ratio falls to 0.6, and approximately 10 percent of the female population has a ratio of 0.4 or less. This low ratio contributes to a persistent elevation (above baseline) of the parathyroid hormone (PTH), the major regulatory hormone affecting Ca and P metabolism.

Figure IV.B.5.2a illustrates the patterns of P and Ca consumption of females in the United States (USDA 1994). These median (50th percentile) intakes are compared to the Recommended Dietary Allowances (RDAs) (National Research Council 1989). Figure IV.B.5.2b shows the Ca:P ratio from the diet of females across the life cycle. The Ca:P ratio of 1:1 is illustrated as a line to demonstrate that this idealized ratio is not met from foods by females (or males) except during the first few months of life when breast milk is consumed.

Requirements and Allowances of Phosphorus

P is an essential nutrient because it is needed for both organic molecules and the mineralized tissues – bones and teeth. Indeed, it has been speculated that calcium-phosphate salts were the substratum for the synthetic steps that resulted in the origin of life in the liquid medium during the early history of the planet Earth.

Phosphorus Requirements

The amounts of P needed in the diet each day depend on several variables, such as stage of development in the life cycle, gender, body size, and physical exertion. Mean requirements of dietary P are not precisely known for either sex during adulthood, but daily U.S.

Median Phosphorus and Calcium Intakes by Females (USDA, 1994)

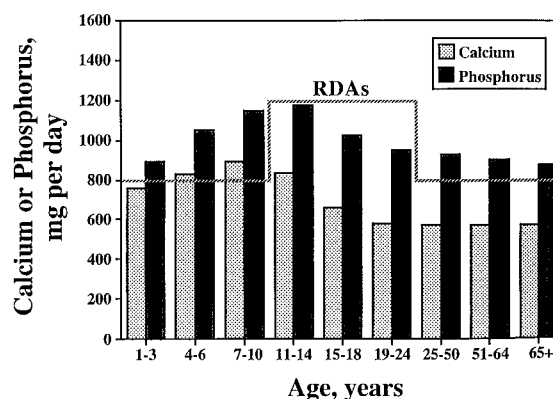


Figure IV.B.5.2a. Median phosphorus and calcium consumption (50th percentile) of females in the United States in comparison to the 1989 Recommended Dietary Allowances (RDAs) for both phosphorus and calcium. (From USDA 1994.)

Calcium to Phosphorus Ratios in the Diets of Females (USDA, 1994)

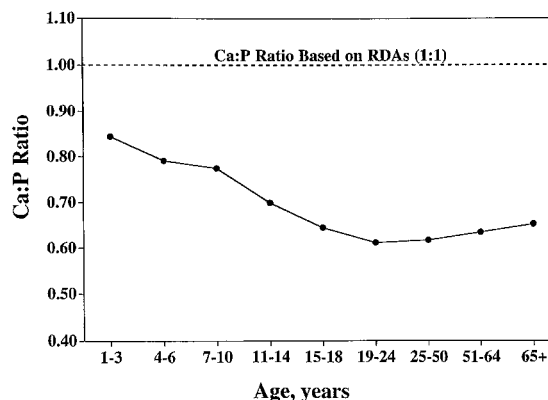


Figure IV.B.5.2b. The dietary calcium to phosphorus (Ca:P) ratio of females across the life cycle, calculated from the data in Figure IV.B.5.2a. The RDAs of calcium and phosphorus generate a 1:1 ratio (dotted line), but this ratio is not achieved except during the first few months of life by infants fed breast milk. (From USDA 1994.)

intakes (roughly between the 10th and 90th percentiles) of P fall in a range of approximately 1,600 to 2,400 milligrams (mg) per day for males and 1,200 to 1,600 mg for females, according to data generated by the U.S. Department of Agriculture's Continuing Survey of Food Intakes by Individuals (CSFII) (USDA 1994; Calvo and Park 1996).

Requirements of P may be as low as 600 to 800 mg a day for females and 800 to 1,000 mg for males, but

these are educated guesses only, which assume adequate consumption patterns of calcium. Excess P_i that is absorbed is excreted by individuals with healthy kidneys in practically a 1:1 ratio to intake. In late life and in individuals who have declining renal function, some phosphate ions may be harbored (not truly stored) in mineralized atheromatous deposits within the arteries and in the skin. P balance-assessment methods have been historically helpful in arriving at estimates of P requirements, but they typically have low precision of measurements in feces; therefore, they are not very reliable.

Phosphorus Allowances

Recommended Dietary Allowances have long been established for phosphorus as well as for calcium, and age- and gender-specific recommendations have been kept identical for both of these minerals across the life cycle since the first edition of the RDAs in the early 1940s. These identical values presume a 1:1 ratio of dietary intake of calcium and total phosphorus (P_i and P_o combined), but actual dietary intakes almost never achieve such a ratio except in the early months or years of life. When the typical intake ratio declines to approximately 0.5:1, the homeostatic regulation of serum calcium becomes so significantly challenged that skeletal mass may be lost in order to maintain the blood calcium concentration at a set level.

The Calcium:Phosphorus Ratio and Relationship

The problem with excessive dietary P intakes (or even just adequate intakes) is the imbalance between calcium and phosphorus that can result from typically low dietary calcium consumption patterns (see Figure IV.B.5.2b). Because practically all foods contain phosphates, but only a few have much calcium, eating behaviors that exclude calcium-rich foods (mainly milk and related dairy foods) may contribute to a condition known as nutritional secondary hyperparathyroidism – and one that can be exacerbated in individuals who consume diets rich in phosphate additives and cola drinks with phosphoric acid (Calvo 1993). The high intake of total P is not in itself so much of a problem as are the behaviors that lead to the avoidance of calcium-rich foods. This is because the latter lowers the Ca:P ratio and causes the development of a persistent elevation of PTH (Calvo, Kumar, and Heath 1990). A diet containing adequate amounts of Ca can overcome these adverse effects of P (Barger-Lux and Heaney 1993).

Because long-term prospective human studies of high-P diets have not been published, data from short-term and cross-sectional investigations must be used to assess the adverse effects of persistent low-Ca, high-P intakes on bone status. Unfortunately, only a few researchers have examined this issue of the dietary Ca:P ratio. One four-week investigation of young women consuming a Ca:P ratio of 0.25:1 revealed the undesirable persistent rise in PTH (Calvo

et al. 1990). Another study of cross-sectional data from healthy, young-adult females indicated that too much P relative to calcium in the usual diet has a negative effect on the mineral content and density of bone (Metz, Anderson, and Gallagher 1993).

The mean Ca:P ratio of U.S. adults approximates 0.5:1 (Calvo 1993), and in fact, healthy ratios of intakes of the two elements range from 0.70 to 0.75 when the recommended number of servings from all food groups (based on the Basic Food Guide or Food Pyramid) are consumed each day. In other words, it is very difficult in the United States to achieve a ratio of 1:1, the ratio recommended in the RDAs (National Research Council 1989), without taking Ca supplements. Nonetheless, a healthy eating pattern should include a ratio within the range of 0.7:1 to 1:1. Intake ratios at or below 0.5:1 are of concern because of the likelihood of persistently elevated PTH concentrations and the potential loss of bone mass, which could lead to fragility fractures.

Digestion of Phosphates from Foods

Phosphates in foods exist mainly as organic molecules that must be digested in order to release inorganic phosphate into fluids of the intestinal lumen. P_i anions freed up by digestive enzymes are then ready for absorption; little likelihood exists for their resynthesis or precipitation because of the lower pH level (–6.0 to 7.0) of the upper half of the small intestine.

The common types of enzymes of the gut that break the bonds of phosphate-containing molecules are secreted almost entirely by the exocrine pancreas. These enzymes include phospholipases, phosphatase, and nucleotidase – as these names imply, the enzymes have specific target molecules that contain phosphates. Once these P_i molecules are solubilized in the lumen, all are equal in the sense that the absorbing mechanisms of the small intestine do not discriminate based on the molecule of origin.

Phytates, which occur in large amounts in cereal grains, are rather poorly digested by humans. The reason is that phytase enzymes are not made in the human body, and only the phytase enzymes present in the bran and other parts of the grain can accomplish this chemical digestion.

Intestinal Absorption of Phosphates

Inorganic phosphate anions are efficiently absorbed across the small intestine, primarily in cotransport with cations, in order to maintain the electrical neutrality of cells. The efficiency of absorption (net) of P_i ranges between 60 and 70 percent in adults, almost twice the efficiency (net) of calcium from the diet (that is, 28 to 30 percent) (Anderson 1991). For example, for every 1,000 moles of P_i , approximately 700 are absorbed (net), compared to only 300 (net) from 1,000 of Ca in the diet. Therefore, the excretory mech-

anisms have to work more efficiently to eliminate the extra P_i absorbed following meals. In children, the net absorption efficiencies may be as high as 90 percent for P_i and 50 percent for Ca. Several factors have adverse effects on P_i absorption, but these typically have little overall influence on P_i utilization, homeostasis, or metabolism. A few factors may enhance P_i absorption, but these, too, have little significance for the overall economy of P_i in the body (Allen and Wood 1994). A high P intake has been reported to have little effect on calcium absorption (Spencer et al. 1978).

The absorption efficiency of P_i declines later in life so that the net absorption of phosphorus from foods is somewhat reduced; probably this occurs in a similar fashion as calcium absorptive efficiency is lowered with age, especially after age 65 in females.

P_i ions are absorbed across all three segments of the small intestine, but the rapid entry into the bloodstream of radioactively labeled phosphates suggests that duodenal absorption occurs both very efficiently and at a high transfer rate. Therefore, the bulk of the absorbed P_i ions are transported across this segment, lesser amounts across the jejunum, and still lesser amounts across the ileum. If the hormonal form of vitamin D – calcitriol or 1,25-dihydroxyvitamin D – is elevated, P_i absorption can be even further enhanced in all segments of the small intestine.

Regulation of Blood Phosphate (P_i) Concentration

The homeostatic regulation of P_i in blood is primarily controlled by PTH, but several other hormones also exert influences on it. The major sites of regulation are the kidneys and the gut. PTH acts on the renal tubules to inhibit P_i reabsorption, while at the same time enhancing Ca reabsorption. The response of the kidneys to the action of PTH is the primary route of loss of P_i from the body. A secondary route is the small intestine, through which intestinal secretions from glands within the serosa of the intestinal lining remove P_i from the bloodstream to the gut lumen. This route of loss of P_i ions is called endogenous fecal excretion, and the quantity of P_i lost by this route may be almost as great as renal losses over a 24-hour period.

Absorption of an excess of P_i tends to lower the blood Ca ion concentration, which then triggers the secretion of PTH from the parathyroid glands. The role of circulating P_i in the secretion of PTH has been investigated in animal models and in human subjects to establish the connection with high dietary phosphorus intake. As mentioned, the absorption of P_i ions is rapid following a meal, much more so than for calcium ions. When an excess of P_i ions exists in the blood, the P_i concentration increases; this change, in turn, drives down the Ca ion concentration through a mechanism involving ionic binding between the two ions.

The net reduction in the concentration of Ca ions then stimulates the secretion of PTH. In turn, PTH enables the transfer of residual circulating Ca and P_i ions into the bone fluid compartment and into other extravascular compartments of the body. (Some investigators suggest that calcitonin, another calcium-regulating hormone, is involved in the movement of these ions into bone and, hence, in the conservation of calcium after a meal.) The persistently elevated PTH, however, tends to undo calcium conservation in the skeleton because this hormone continuously stimulates the reverse transfer of Ca and P_i ions from bone to blood (Calvo et al. 1990). The net result is that the skeleton loses bone mineral when PTH is elevated, even within the normal range of blood concentration, over extended periods of time. (The actual site of loss of this Ca is the gut, which has a poorly regulated secretion of Ca ions through intestinal glands.) PTH also has other roles in the kidney that, in effect, contribute to Ca retention by the body and to the elimination of P_i via urinary excretion.

PTH is considered the major hormone regulating P_i homeostasis because of its powerful roles in enhancing renal and, possibly, intestinal P_i losses while, at the same time, conserving calcium ions. When PTH is elevated, renal P_i reabsorption is largely inhibited, and similarly, the secretion of P_i ions by intestinal mechanisms is enhanced (although an understanding of this route of P_i loss is less established). PTH also acts on bone tissue to increase the transfer of calcium ions from the bone fluid compartment (BFC) and from the resorption of mineralized bone tissue to the blood plasma to restore the calcium ion concentration. By these same actions of PTH, P_i ions are also indirectly transferred from the BFC and bone to the blood.

Phosphate Balance

P balance means that intake of P from foods equals losses in urine and feces (and other sources, such as sweat and skin, which are seldom measured). In effect, P_i ions that are absorbed are accounted for by losses from the body. Under balance conditions, no net gain or loss of P_i ions occurs. This zero-balance state probably only exists during the adult years from roughly 20 to 60. During growth, and during pregnancy and lactation, positive balance states tend to predominate, whereas in late life, phosphate retention may increase and become a major health problem for individuals with declining renal function. Phosphate retention (positive P balance) results from the declining effectiveness of PTH in enhancing renal excretion of P_i ions with decreasing renal function.

A schematic diagram of the P balance of an adult male is shown in Figure IV.B.5.3. An adult male would typically consume 1,200 to 1,400 mg of P a day, whereas an adult female would consume 900 to 1,000 mg per day (USDA 1994; Calvo and Park 1996).

These estimated intakes of P by gender, however, do not include phosphate additives in foods.

Positive P balance (both P_o and P_i), or the net gain of this element by the body, is difficult to measure, but numerous balance studies suggest that P homeostasis (that is, zero balance) is typically maintained even when Ca balance may be significantly negative. Radioisotopic and stable nuclide studies have greatly advanced our knowledge about the fluxes of P_i and Ca ions across the gut and renal tubules in animal models and, to a lesser extent, in human subjects. Balance studies without the use of stable or radioactive nuclides are notoriously fraught with potential errors of collection and measurement, and these difficulties make such studies generally unreliable in the precise quantitative sense.

The uptake of P_i ions by cells requires carrier mechanisms or cotransport systems because of the electrical charge and water-solubility properties of these anions. P_i ions typically cotransport with glucose in postprandial periods, but their charge must be neutralized by cations, typically not calcium ions. Also, after meals, P_i ions enter the bone fluid compartment, but in this case, typically with calcium ions. Calcitonin has been considered primarily responsible for the uptake by bone tissue of these two ionic species following food ingestion and the intestinal absorption of the ions (Talmage, Cooper, and Toverud 1983). After entry into cells, the P_i ions in the cytosol are almost immediately used to phosphorylate glucose or other molecules, and a small fraction of the ions are stored as organic molecules or inorganic salts within cellular organelles.

Some P_i ions that enter bone may enter bone cells,

especially osteoblasts or lining cells, whereas other ions bypass the cells and go directly to the BFC, an extension of the blood/extracellular-fluid continuum. In the BFC, P_i ions in solution increase the P_i concentration (activity) that permits these ions to combine with Ca ions in excess of their solubility product constant (K_{sp}) and form mineral salts (precipitate) in bone extracellular tissue. The formation of hydroxyapatite crystals (mineralization) is essential for structural support and protection of internal organs from environmental trauma. P_i ions are, therefore, essential for the formation of the endoskeletons typical of most vertebrates except cartilaginous fish.

Approximately 60 to 70 percent of P_i ions are cleared by the kidneys in healthy individuals. If PTH is elevated, P_i excretion is enhanced even more, so that P_i losses are further increased. Under the same conditions of elevated PTH, the secretion of P_i by the gut is also increased. The endogenous fecal secretion of phosphates is the second major route of loss that the body uses to maintain P_i ion homeostasis.

Persistently Elevated Parathyroid Hormone

Of the many diseases that have significant alterations in P homeostasis, only two are reviewed in any depth here. The first is persistently elevated PTH in response to a low Ca:P dietary intake pattern, whereas the second is renal secondary hyperparathyroidism resulting from chronic renal failure. Previously, the former was often referred to – perhaps erroneously in the case of humans (in contrast to animals) – as nutritional secondary hyperparathyroidism. In the context of human disease, the term “hyperparathyroidism” is inappropriate.

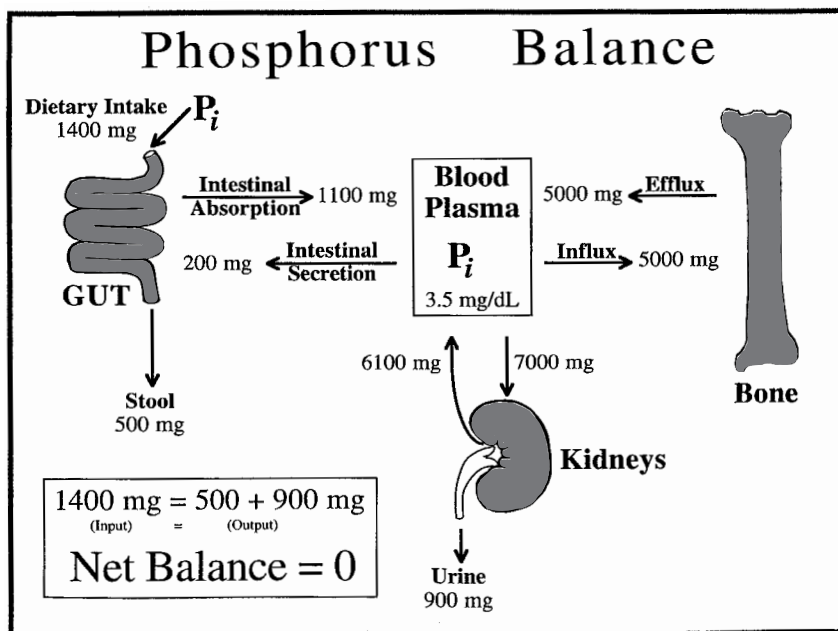


Figure IV.B.5.3. Schematic diagram of phosphorus balance of an adult male. (Adapted from various sources.)

ate because the PTH levels that result from a low Ca:P ratio typically remain within the normal range of blood concentration, although usually at the high end of the range. As mentioned, a persistently elevated PTH, even if it remains within the normal range, contributes to increased bone turnover that can result in a reduction of bone mass and density (Calvo et al. 1990). If this condition continues for a year or longer, it could contribute to fragility fractures because of the thinning of trabecular plates at bone sites, such as the vertebrae, wrist, and proximal femur. On the basis of obtaining a benefit from a PTH value at the lower end of the range, individuals with a low Ca:P ratio would be advised to increase their calcium intake from foods first and from supplements second. An adequate calcium intake is known to reduce serum PTH concentration (Krall and Dawson-Hughes 1994).

Figure IV.B.5.4 diagrams the mechanism through which a low dietary Ca:P ratio contributes to the development of a persistently elevated PTH concen-

tration. Figure IV.B.5.5 illustrates the potential changes in bone mass and mineralization of the skeleton in individuals who typically consume diets with low Ca:P ratios compared to those who have normal intake ratios. The persistently elevated PTH is responsible for the limited bone mineralization and the loss of bone mass (Anderson 1996).

Renal Secondary Hyperparathyroidism

Renal secondary hyperparathyroidism results from a severe increase in PTH that occurs because the kidneys can no longer filter and secrete sufficient amounts of P_i ions each day. As the blood concentration of P_i increases, the serum PTH also rises in an attempt to correct the error (increase of serum P_i). The action of PTH on bone tissue then predominates, and the rate of bone turnover continues to increase, unless corrected by renal dialysis or kidney transplantation. Without correction, the net result is a

Persistent Elevation of Parathyroid Hormone

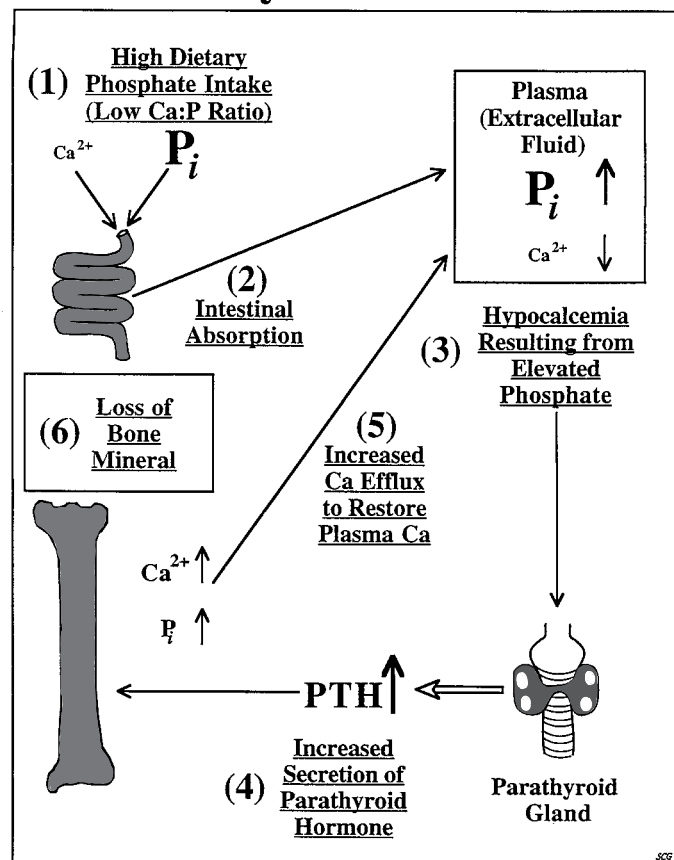


Figure IV.B.5.4. Mechanism through which a low dietary calcium:phosphorus ratio contributes to the development of a persistently elevated parathyroid hormone (PTH) concentration in the blood.

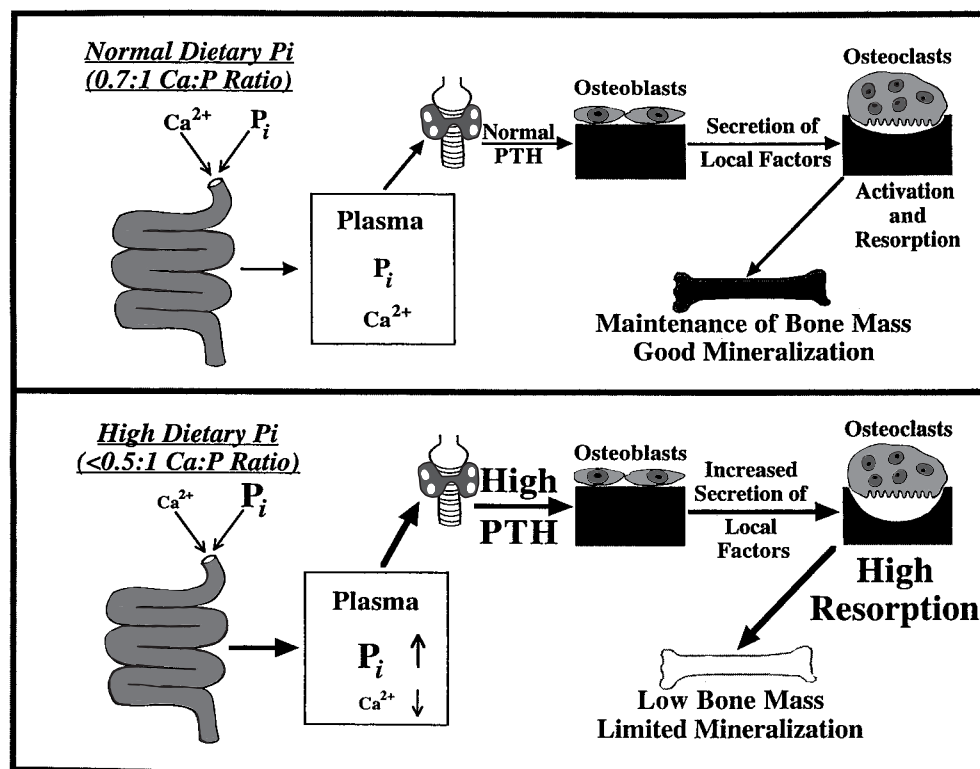


Figure IV.B.5.5. Comparison of parathyroid hormone (PTH) responses of normal and high dietary phosphorus (P_i) and effects of PTH on bone mass.

continuing increase in the serum P_i concentration and a rapid thinning of bone tissue at practically every site in the body. If severe enough, this condition can result in fractures at almost any skeletal location. Oral phosphate binders, such as aluminum or magnesium hydroxides, are usually administered to patients to reduce the amount of P_i absorbed by the small intestine and to enhance calcium absorption, but this strategy typically is not sufficient to stem the gradual increase in serum PTH as the disease progresses.

Potential Adverse Effects

“Over-the-counter” antacids contain mineral salts, which bind phosphate ions that are released from foods or secreted into the gut lumen by glands of the GI tract. If excessive amounts of these antacid drugs (also taken as nutrient supplements when containing calcium and/or magnesium) are ingested, individuals may be at increased risk of lowering their serum phosphate concentration to the point of serious deficiency. It is fortunate that the use of mineral antacids containing calcium, magnesium, aluminum, or some combination has declined greatly of late in the United States because of the availability of more

effective antacid drugs that previously could only be obtained by prescription.

Summary

Phosphorus has important roles in human health. Phosphates participate in diverse functions in the body, both intracellularly and extracellularly as P_i and P_o groups. These anions are especially important in energetic reactions within cells, in nucleic acids, and in other structural molecules and the extracellular tissues of bones and teeth. Dietary deficiency of this element is highly unlikely during practically the entire life cycle.

High intakes of P are common because of the natural widespread availability of phosphates in foods. Moreover, processed foods are likely to contain phosphate additives that contribute to a potentially excessive consumption of phosphorus by individuals in the United States and, most likely, in all economically developed nations throughout the world. When a chronic pattern of low Ca consumption is coupled with high dietary intakes of P (both P_o and P_i), PTH becomes persistently elevated. The potential outcome is low bone mass and an increased risk of skeletal fractures, especially late in life. This adverse relationship between a high dietary P intake and bone loss

strongly suggests that the consumption of adequate amounts of Ca are essential for the development and maintenance of bone mass throughout life.

John J. B. Anderson

Bibliography

- Allen, L. H., and R. J. Wood. 1994. Calcium and phosphorus. In *Modern nutrition in health and disease*. Eighth edition, ed. M. E. Shils, J. A. Olson, and M. Shike, 144–63. Philadelphia, Pa.
- Anderson, J. J. B. 1991. Nutritional biochemistry of calcium and phosphorus. *Journal of Nutritional Biochemistry* 2: 300–9.
1996. Calcium, phosphorus and human bone development. *Journal of Nutrition* 126: 1153S–8S.
- Anderson, J. J. B., and C. J. H. Barrett. 1994. Dietary phosphorus: Benefits and problems. *Nutrition Today* 29: 29–34.
- Anderson, J. J. B., and S. C. Garner, eds. 1966. *Calcium and phosphorus in health and disease*. Boca Raton, Fla.
- Barger-Lux, M. J., and R. P. Heaney. 1993. Effects of calcium restriction on metabolic characteristics of premenopausal women. *Journal of Clinical Endocrinology and Metabolism* 70: 264–70.
- Bronner, F., and J. W. Coburn, eds. 1981. *Disorders of mineral metabolism*. 3 vols. New York.
- Burckhardt, P., and R. P. Heaney, eds. 1994. *Nutritional aspects of osteoporosis, II*. Sero Symposium. Rome.
- Calvo, M. S. 1993. Dietary phosphorus, calcium metabolism, and bone. *Journal of Nutrition* 123: 1627–33.
- Calvo, M. S., R. Kumar, and H. Heath III. 1990. Persistently elevated parathyroid hormone secretion and action in young women after four weeks of ingesting high phosphorus, low calcium diets. *Journal of Clinical Endocrinology and Metabolism* 70: 1340–4.
- Calvo, M. S., and Y. K. Park. 1996. Changing phosphorus content of the U.S. diet: Potential for adverse effects on bone. *Journal of Nutrition* 126: 1168S–80S.
- Dziezak, J. D. 1990. Phosphates improve many foods. *Food Technology* 44: 80–92.
- Favus, M. J., ed. 1996. *Primer on the metabolic bone diseases and disorders of mineral metabolism*. Third edition. Philadelphia, Pa.
- Hands, E. S. 1990. *FOOD finder: Food sources of vitamins and minerals*. Second edition. Salem, Ore.
- Krall, E. A., and B. Dawson-Hughes. 1994. Osteoporosis. In *Modern nutrition in health and disease*. Eighth edition, ed. M. E. Shils, J. A. Olson, and M. Shike, 1559–68. Philadelphia, Pa.
- Lobaugh, B. 1996. Blood calcium and phosphorus regulation. In *Calcium and phosphorus in health and disease*, ed. J. J. B. Anderson and S. C. Garner, 27–43. Boca Raton, Fla.
- Metz, J., J. J. B. Anderson, and P. N. Gallagher, Jr. 1993. Intakes of calcium, phosphorus, protein and level of physical activity are related to radial bone mass in young adult women. *American Journal of Clinical Nutrition* 58: 537–42.
- National Research Council, Subcommittee on the Tenth Edition of the RDAs, Food and Nutrition Board, Commission on Life Sciences. 1989. *Recommended dietary allowances*. Tenth edition. Washington, D.C.
- Nordin, B. E. C., ed. 1976. *Calcium, phosphorus and magnesium*. London.
1988. *Calcium*. New York.
- Oenning, L. L., J. Vogel, and M. S. Calvo. 1988. Accuracy of methods estimating calcium and phosphorus intake in daily diets. *Journal of the American Dietetic Association* 88: 1076–81.
- Spencer, H., L. Kramer, D. Osis, and C. Norris. 1978. Effect of phosphorus on the absorption of calcium and on the calcium balance in man. *Journal of Nutrition* 108: 447–53.
- Talmage, R. V., C. W. Cooper, and S. U. Toverud. 1983. The physiologic significance of calcitonin. In *Bone and mineral research annual*, Vol. 1, ed. W. A. Peck, 74–143. Amsterdam.
- USDA (U.S. Department of Agriculture). 1994. *Continuing survey of food intakes of individuals (CSFII): Diet and health knowledge survey 1991*. Springfield, Va.

IV.B.6 Potassium

Potassium (K) is found in virtually all aerobic cells and is essential to life. It is the third most abundant element in the human body (after calcium and phosphorus) and the eighth most abundant element in the earth's crust, with a mass percent of 1.8, which means that every 100 grams (g) of the earth's crust contains 1.8 g of potassium. Potassium is a very reactive alkali metal with an atomic number of 19 and an atomic weight of 39.098 atomic mass units (amu). Its outer "4s" electron is not bound very tightly to the atom, which is therefore easily ionized to K⁺ (Dean 1985), and potassium reacts readily with chlorine to form the salt potassium chloride. Potassium chloride is a white crystalline solid at room temperature with alternating potassium ions and chloride ions on the lattice sites. Potassium is found primarily in seawater and in natural brines in the form of chloride salt. The minerals mica and feldspar also contain significant quantities of potassium (Dean 1985).

The Discovery of Elemental Potassium

Potassium was first isolated in 1807 by Humphry Davy (1778–1829), who electrolyzed "potash" with a newly invented battery designed to contain a series of voltaic cells, with electrodes made out of zinc and copper plates dipped in a solution of nitrous acid and alum. In Davy's time, the term "potash" referred to any number of different compounds, including "vitriol of potash" (potassium sulfate), "caustic potash" (potassium hydroxide), and "muriate of potash" (potassium chloride as well as potassium carbonate), the last of which was formed by leaching ashes from a wood fire and evaporating the solution to near dryness in an iron pot. Today, potash is usually potassium carbonate, although potassium chloride is still called potash by fertilizer manufacturers (Kent 1983: 262). The potash Davy used was potassium

hydroxide that he had dried and melted. He wrote of his experiment:

A small piece of pure potash, which had been exposed for a few seconds to the atmosphere, so as to give conducting power to the surface, was placed upon an insulated disc of platina, connected with the negative side of a battery of the power 250 of 6 and 4, in a state of intense activity; and a platina wire communicating with the positive side, was brought in contact with the upper surface of the alkali. The whole apparatus was in the open atmosphere. Under these circumstances a vivid action was soon observed to take place. The potash began to fuse at both its points of electrization. There was a violent effervescence at the upper surface; at the lower, or negative surface, there was no liberation of elastic fluid; but small globules having a high metallic lustre, and being precisely similar in visible characters to quicksilver formed, and others remained and were merely tarnished, and finally covered with a white film which formed on their surfaces. These globules, numerous experiments soon shewed to be the substance I was in search of, and a peculiar inflammable principle the basis of potash. (Davy 1839–40, 5: 60)

The next day, Davy isolated sodium metal by electrolyzing soda ash (sodium hydroxide) in much the same way. In the history of chemistry, isolating potassium and sodium was no mean accomplishment. It had been suspected by several people, but especially by Antoine Lavoisier (1743–94), that potash was a compound and that the “basis of potash” was, indeed, a metal (Partington 1962, 3: 485). Davy’s experiments confirmed this suspicion. Several years earlier, in 1801, Carl Friedrich Kielmeyer, also suspecting that potash was an oxide of some metal, had attempted to electrolyze potash using a voltaic pile but was unsuccessful (Partington 1964, 4: 45). Thus, the credit for discovering the two most important alkali metals clearly goes to Humphry Davy.

Between 1808 and 1809, the French chemists Louis Thenard (1777–1857) and Joseph Gay-Lussac (1778–1850) found that only small quantities of potassium and sodium could be derived by the electrolysis of fused alkali hydroxides and went on to develop a much improved method for producing larger quantities of both (Partington 1964, 4: 94). Thenard and Gay-Lussac reacted the fused alkali with red-hot iron turnings in an iron gun barrel lined with clay and sand and collected the condensed metal vapor in a receiver attached to the gun barrel. An explosion using this dangerous device nearly blinded Gay-Lussac.

A variation that further improved the method for producing potassium used potassium carbonate as the source of potassium and carbon instead of iron as the reducing agent to produce elemental potassium and carbon dioxide as the reaction products. In 1827,

Frederich Wohler (1800–82) first employed potassium produced by this technique to isolate metallic aluminum in more or less pure form. He reacted anhydrous aluminum chloride with potassium metal as the reducing agent and obtained enough aluminum metal to measure its properties (Ihde 1964: 467). Today, metallic potassium is usually produced by reacting molten sodium with molten potassium chloride and condensing the gaseous potassium formed by this reaction. There are very few industrial uses for elemental potassium, although many of its compounds are widely utilized throughout industry and agriculture. For example, potassium nitrate is commonly employed as a fertilizer in the tobacco industry where chloride-containing fertilizers are undesirable.

Potassium in Living Organisms

Despite all of potassium’s various functions, its principal function in living organisms is in the transportation of ions across cell membranes. In most animal cells, the internal concentration of potassium ions is between 20 and 30 times higher than the external concentration of potassium ions found in the extracellular fluids. Most cells also have considerably different internal concentrations of sodium ions than are found externally. Neither the reason for the existence of these ionic gradients across the cell membrane nor, for that matter, potassium’s principal role in cell metabolism was understood until comparatively recently, and many of potassium’s functions in living organisms are still being investigated. It can be said with certainty, however, that moving sodium ions and potassium ions across membranes is an important activity in most organisms, and that if this activity stops, the organism dies.

The Sodium-Potassium-ATPase Pump

The significant step toward understanding potassium’s role in animal cells was taken in 1957. In that year, the Danish biochemist Jens Skou, who later won the 1997 Nobel Prize in chemistry, found an enzyme in crab nerve cells that hydrolyzed adenosine triphosphate (ATP) into adenosine diphosphate (ADP) and any one of several phosphorus-containing anions, such as dihydrogen phosphate, in the presence of magnesium ions only if both sodium ions (Na^+) and potassium ions (K^+) were present (Skou 1965: 6). Magnesium ions are always required for enzymes (called ATPases) to catalyze the hydrolysis of ATP.

The unusual property of this enzyme was that neither K^+ nor Na^+ alone had any significant effect on its activity. A short time later, Skou proposed that this $\text{Na}^+\text{-K}^+\text{-ATPase}$ complex was part of a transmembrane pump that pumped Na^+ and K^+ into and out of cells, and that the energy needed for this process was supplied by the hydrolysis of ATP. Many different kinds of animal tissues were soon found to exhibit similar $\text{Na}^+\text{-K}^+\text{-ATPase}$ activity, and it was shown that this

enzyme was, indeed, a protein that resided in a cell's membrane, with its sodium ion receptor facing the interior of the cell (or the cytoplasm), and its potassium ion receptor facing the external environment. The $\text{Na}^+\text{-K}^+\text{-ATPase}$ pumped sodium ions out of the cell and potassium ions into the cell, accompanied by the hydrolysis of ATP. Brain, nerve, and muscle cells – and the electric organ of the electric eel – were discovered to be particularly rich in $\text{K}^+\text{-Na}^+\text{-ATPase}$ activity (Lehninger 1970:617).

“Membrane transport” – the pumping of ions across cell membranes – is such an important part of a cell's total activity that it is estimated that more than one-third of the ATP consumed by a resting animal is expended transporting ions across membranes (Stryer 1988: 950). The $\text{Na}^+\text{-K}^+$ pump, in particular, maintains proper electrochemical potentials across cell membranes and maintains the proper concentrations of Na^+ and K^+ , both internally and externally. Because these concentrations are usually far from the equilibrium concentrations, the ions must be pumped against their respective concentration gradients. The sodium-ion and potassium-ion concentration gradients in most animal cells control cell volume, drive the transport of sugars and amino acids across cell membranes, and control the electrical excitability of both nerve and muscle cells.

In the most ubiquitous transport systems in animal cells, the $\text{Na}^+\text{-K}^+$ pump removes 3 sodium ions from the interior of the cell and pumps in 2 potassium ions from the cell's surroundings, using the transmembrane $\text{Na}^+\text{-K}^+\text{-ATPase}$ to facilitate the transport of the ions. One molecule of ATP is hydrolyzed to supply the necessary free energy. This process is shown schematically in Figure IV.B.6.1. Typically, the internal concentration of potassium ions is approximately 20 times greater than the external concentration of potassium ions. This concentration gradient is maintained also by the free energy supplied by the hydrolysis of ATP and the transmembrane enzyme pump. The ATP supplies

energy to the enzyme by phosphorylating it in the presence of sodium ions and magnesium ions. The phosphorylated enzyme is then dephosphorylated in the presence of potassium ions to regenerate the original enzyme and form a phosphorus-containing ion, such as hydrogen phosphate. The energy stored in the ATP is transferred to the enzyme via the phosphate group bonding to it. The enzyme uses this energy to move ions against their gradients and then gets rid of the phosphate group via hydrolysis.

Potassium in the Nerves and the Nervous System

Potassium ions and sodium ions are found in the nerve cells of virtually all animals and play a vital role in the transmission of nerve impulses. These ions regulate the nerve's transmembrane potential, and it is the transmembrane potential difference between the interior and exterior of a nerve cell that causes the transmission of nerve impulses along the nerve (Atkins 1994: 334). Many $\text{K}^+\text{-Na}^+$ pumps are distributed throughout the nervous system. When a nerve is resting, there is a high internal potassium ion concentration and a high external sodium ion concentration. When the cell experiences a pulse, the nerve cell membrane's structure alters, becoming permeable to sodium ions. Because the membrane is now more permeable to sodium ions, the ions rapidly flow into the interior of the cell and reduce the size of the sodium ion concentration gradient so that it becomes smaller than it was when the nerve was resting. In other words, the Na^+ ions and K^+ ions exchange places. Because the membrane's potential arises primarily from the difference in concentration on either side of the membrane, the voltage across the membrane drops. This change in voltage triggers the adjacent part of the cell wall to alter its structure. The pulses of collapsing potential pass along the nerve. Behind each pulse, the $\text{Na}^+\text{-K}^+$ pump restores the proper internal sodium ion concentration by pumping out sodium ions and pumping in potassium ions.

Potassium in Muscle Tissue

Generally speaking, muscle tissue can be divided into two classes, smooth and striated. Striated muscle is under voluntary control and has a striated or striped appearance under a light microscope. Striated muscle found in vertebrates contains 2 protein filaments that interact with each other. One of the filaments contains myosin and the other contains 3 proteins – actin, tropomyosin, and troponin. Striated muscle contraction is regulated primarily by calcium ions (Ca^{++}), and the calcium ion concentration is itself regulated by a $\text{Ca}^{++}\text{-ATPase}$ pump, very similar in kind and function to the $\text{Na}^+\text{-K}^+\text{-ATPase}$ pumps found in other kinds of cells. In mammalian striated muscle, however, fatigue is associated with a loss of intercellular K^+ and a gain in intercellular Cl^- , Na^+ , and H_2O , and it may indeed be the case that, in humans, fatigue is related to these changes in the potassium ion concentration

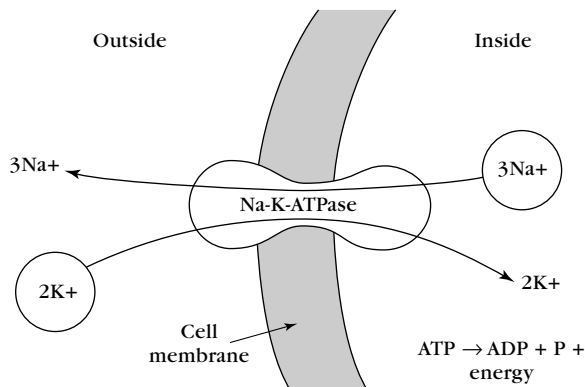


Figure IV.B.6.1. The $\text{Na}^+\text{-K}^+\text{-ATPase}$ transmembrane pump pumping Na^+ ions out of the cell and K^+ ions into the cell.

gradient across muscle cells. Lowering the internal potassium ion concentration and raising the external potassium ion concentration depolarizes the cell membrane. The rapid recovery of muscle function following brief rest periods is caused by a reestablishment of the proper potassium ion concentrations and a restoration of the resting potential of the muscle cell's membrane.

Smooth muscle, such as that found in the heart or veins, is not subject to voluntary control. The contraction of smooth muscle is controlled by the degree of phosphorylation of its light chains. Phosphorylation leads to contraction and dephosphorylation leads to relaxation. In smooth muscle, $\text{Na}^+\text{-K}^+$ pumps and active transport are directly involved with contraction and relaxation. For example, in vascular smooth muscle, the function of the $\text{Na}^+\text{-K}^+\text{-ATPase}$ pump is to transport 2 potassium ions into the cell for every 3 sodium ions it takes out. The energy for this transport is supplied by phosphorylation. As in striated muscles, extracellular calcium ions also play an important role in cell function, but in smooth muscles this role is primarily to regulate the $\text{K}^+\text{-Na}^+$ equilibria extant in the cell (O'Donnell and Owen 1994).

The importance of the $\text{Na}^+\text{-K}^+$ pump in smooth muscle function has some intriguing consequences. It is entirely possible – and evidence is mounting that it is more than just possible – that a principal underlying cause of hypertension in humans is an inhibitor of $\text{Na}^+\text{-K}^+\text{-ATPase}$ activity or an impaired $\text{Na}^+\text{-K}^+$ pump (Blaustein et al. 1986; O'Donnell and Owen 1994: 687). It is also well known that a certain class of compounds, called cardiotonic steroids, specifically inhibit the $\text{Na}^+\text{-K}^+$ pump. These steroids prevent the dephosphorylation of $\text{Na}^+\text{-K}^+\text{-ATPase}$. As a consequence, these compounds are very important in the treatment of heart disease.

For example, digitalis (an extract from purple foxglove leaves that contains a mixture of cardiac steroids), which is perhaps the best-known remedy for congestive heart failure and has been widely used for centuries, increases the force of heart muscle contraction by inhibiting the $\text{Na}^+\text{-K}^+\text{-ATPase}$ pump (Voet and Voet 1990: 496). This increases the sodium ion concentration inside the cell, thereby causing a reduction in the concentration gradient across the membrane, because the internal sodium ion concentration becomes closer to the external sodium ion concentration. Reducing the sodium ion concentration gradient reduces the extrusion of calcium ions via the $\text{Na}^+\text{-Ca}^{++}$ exchanger, which increases the internal calcium ion concentration. A high intracellular calcium ion concentration causes the heart muscle to contract (Stryer 1988: 955). Ion exchangers, such as the $\text{Na}^+\text{-Ca}^{++}$ exchanger, are conduits – or ports – that allow ions to diffuse in and out of a cell via concentration gradients rather than through active transport, and they are not fueled by ATP.

The medicinal effects of digitalis were known long before the compounds themselves were identified

and long before the $\text{Na}^+\text{-K}^+\text{-ATPase}$ pump was discovered. In the 1770s, a woman in Shropshire, England, using extracts from some 20 different herbs, prepared a cocktail that had a remarkable effect on curing congestive heart failure or “dropsy,” as it was called. Many people suspected that this woman was a witch because of the curative power of her concoctions. A physician by the name of William Withering heard about the woman's remarkable medicine and, after considerable effort, found that foxglove was the significant herb in the cocktail. He published his findings in 1785 in his classic work, *An Account of the Foxglove and Some of its Medical Properties*. A line from this paper is worth noting: “It [the foxglove] has a power over the motion of the heart to a degree yet unobserved in any other medicine, and this power may be converted to salutary ends” (Estes and White 1965: 110).

Potassium in Protein Synthesis

Protein synthesis on ribosomes requires a high potassium concentration for maximum efficiency. Ribosomes are cellular bodies that serve as the sites for protein synthesis and can be thought of as the cell's protein factories because protein assembly from amino acids, controlled by RNA, takes place on the ribosome's surface (Lehninger 1970: 616).

Potassium in Glycolysis

Potassium is necessary for glycolysis, which is a form of fermentation that ultimately converts glucose ($\text{C}_6\text{H}_{12}\text{O}_6$) into pyruvate (C_3COCO^-) with the associated production of ATP. Glycolysis can be thought of as a fundamental aspect of the generation of metabolic energy that occurs in virtually all living organisms, including humans. Glycolysis requires potassium ions for maximum activity of one of the enzyme catalysts, pyruvate kinase, which is involved in the process (Stryer 1988: 350). If there is sufficient oxygen present, the pyruvate enters the mitochondria where it is completely oxidized to CO_2 and H_2O . If there is insufficient oxygen present, as is often the case in muscle contraction, pyruvate is converted to lactic acid. In yeast, which is an anaerobic organism, pyruvate is converted to ethanol and carbon dioxide.

In 1897 and 1898, the German chemist Eduard Buchner discovered that when yeast cells were crushed with sand and diatomaceous earth (*kieselguhr*), the cell-free liquid extract was able to ferment sucrose into alcohol and carbon dioxide. Buchner's results showed, for the first time, that fermentation could occur outside a living cell, and his fermentation experiments contributed greatly to the overthrow of the “vital force” theory then prevalent in the biological sciences, and started the field of modern biochemistry. Even Louis Pasteur thought that fermentation could occur only in living cells. Buchner was awarded the Nobel Prize in chemistry for this work in 1907 (Partington 1964, 4: 309).

Renal Control of Potassium

In humans, the kidneys are responsible for most potassium excretion, although fecal matter does contain about 10 percent of the potassium ingested (Stanton and Giesbisch 1990). However, renal malfunction as a result of disease or trauma often will prevent the proper elimination of potassium from the body. For example, because maintenance of the proper potassium ion balance is vital, the secretion of potassium ions by the distal tubule is one of the kidney's most important functions, and it is the distal tubule that, more or less, regulates the quantity of potassium ions eliminated in the urine. More than 75 percent of the filtered potassium is reabsorbed in the proximal tubule, and this percentage remains nearly constant no matter how much is filtered. About 50 percent of the urinary potassium is secreted into the urine by the distal tubules in normally functioning kidneys. If this secretion is prevented – as, for example, in polycystic kidneys – dangerous levels of potassium ions can accumulate and cause heart failure. $\text{Na}^+\text{-K}^+\text{-ATPase}$ pumps, $\text{H}^+\text{-K}^+\text{-ATPase}$ pumps, and $\text{K}^+\text{-Cl}^-\text{-Na}^+$ cotransport have all been found to control potassium in tissues within the kidneys.

Difficulties with potassium elimination are also encountered if the production of a hormone in the adrenal gland, aldosterone, is inhibited. Aldosterone catalyzes the elimination of potassium ions so that lowering the concentration of this enzyme reduces the rate of potassium ion elimination and can increase potassium ion concentrations in the blood to dangerous levels.

Potassium and Health

Potassium is essential for human life. A normal adult male, weighing 70 kilograms (kg), contains approximately 135 g of the mineral in his body. About 98 percent of this potassium is found in his cells and the other 2 percent in extracellular fluids (Macrae, Robinson, and Sadler 1993, 6: 3668). Potassium deficiency is called hypokalemia and – because potassium is present in most foods – when it occurs, it is often in areas where people exist on subsistence or starvation diets. Severe diarrhea, diabetes, and prolonged use of cortisone can also cause hypokalemia.

For living cells to function properly it is essential they maintain a correct balance between internal potassium ion concentrations and external potassium ion concentrations. There are many factors that influence this balance, in addition to the various pumps and transport systems already described. For example, there must be a normal water balance in the organism, because if the amount of K^+ inside a cell remains fixed and the amount of water increases, the potassium ion concentration will decrease proportionately, or if the amount of water decreases, the potassium ion concentration will increase. A potassium deficiency resulting from urine loss often occurs during the treatment of heart disease because the medication used prevents

sodium and water retention. To reduce this deficiency, foods high in potassium are often prescribed.

Recently, a possible connection between potassium and hypertension has been discovered (Brancati 1996). It was found that African-Americans, a particularly high-risk group for hypertension and acute myocardial infarction, benefited greatly from a diet rich in potassium. Indeed, in a double-blind test, all of the subjects who received potassium supplements reduced their blood pressure – regardless of age, gender, body weight, or alcohol consumption. It was not clear why potassium was especially beneficial for African-Americans. It may be that their diets are particularly low in potassium, but it is more likely the case that, for some reason, African-Americans are especially sensitive to potassium. This finding is certainly also consistent with the mechanism of the $\text{Na}^+\text{-K}^+\text{-ATPase}$ pump discussed earlier.

Potassium and Diet

Potassium is one of the most important elements in the human diet. According to the National Academy of Sciences, a healthy adult should consume between 1,875 milligrams (mg) and 5,625 mg of this mineral daily (National Academy of Sciences 1980: 173). It is probable that the diets of our hunter-gatherer ancestors contained some 2 mg of potassium for every calorie consumed. It is also known that people who consume more than 4 g of potassium each day have a much lower incidence of disease.

Unfortunately, a modern diet typically contains only 0.5 mg of potassium for every calorie consumed, although it is the case that practically all successful dietary weight-loss programs, dietary cholesterol-lowering programs, and dietary blood-pressure-lowering programs contain foods that are high in potassium. Some of these are oranges, tomatoes, peas, spinach, bananas, cantaloupe, and fish (Pennington 1985). Foods that are very high in potassium content, with the exception of potatoes, are relatively low in energy content, which makes them especially healthful for normal adults.

Summary

Potassium is a relatively abundant alkali metal that comprises a significant fraction of the earth's crust. Its salts and oxides are widely used in industry and agriculture. Potassium's role in membrane transport and other metabolic processes make it vital to virtually every living organism. Recently, antibiotics using potassium ions have been discovered, and it is beginning to appear that eating foods rich in potassium or taking potassium supplements can reduce hypertension in humans. Most foods contain some potassium, and the maintenance of good health in normal adults seems to require a diet that contains 4 g of potassium per day.

David S. Newman

Bibliography

- Atkins, P. W. 1994. *Physical chemistry*. Fifth edition. New York.
- Blaustein, M. P., T. Ashida, W. F. Goldman, et al. 1986. Sodium/calcium exchange in vascular smooth muscle: A link between sodium metabolism and hypertension. *Annals of the New York Academy of Science* 448: 199–216.
- Brancati, F. 1996. Effect of potassium supplementation on blood pressure in African Americans consuming a low potassium diet. *Archives of Internal Medicine* 8: 61–7.
- Davy, J. 1839–40. *The collected works of Sir Humphry Davy*. 9 vols. London.
- Dean, J. A. 1985. *Lange's handbook of chemistry*. Thirteenth edition. New York.
- Estes, J. W., and P. D. White. 1965. William Withering and the purple foxglove. *Scientific American* 212: 110–17.
- Fitts, R. H. 1994. Cellular mechanism of muscle fatigue. *Physiological Review* 74: 49–94.
- Ihde, A. J. 1964. *The development of modern chemistry*. New York.
- Kent, J. A. 1983. *Riegel's handbook of industrial chemistry*. New York.
- Lehninger, A. L. 1970. *Biochemistry*. New York.
- Macrae, R., R. K. Robinson, and M. J. Sadler, eds. 1993. *Encyclopedia of food science, food technology and nutrition*. 8 vols. London and San Diego, Calif.
- National Academy of Sciences. 1980. *Recommended dietary allowances*. Ninth edition. Washington, D.C.
- O'Donnell, M. E., and N. E. Owen. 1994. Regulation of ion pumps and carriers in vascular smooth muscle. *Physiological Reviews* 74: 683–720.
- Partington, J. R. 1962–70. *A history of chemistry*. 4 vols. London.
- Pennington, Jean A. 1985. *Bowes and Church's food values of portions commonly used*. New York.
- Petrucchi, R. H. 1989. *General chemistry*. Fifth edition. New York.
- Skou, J. C. 1965. Enzymatic basis for active transport of Na⁺ and K⁺ across the cell membrane. *Physiological Reviews* 45: 496–617.
- Stanton, B., and G. Giesbisch. 1990. *Handbook of physiology*. Second edition. New York.
- Stryer, L. 1988. *Biochemistry*. Third edition. New York.
- Voet, D., and J. G. Voet. 1990. *Biochemistry*. New York.
- Withering, W. 1785. *An account of the foxglove and some of its medical properties*. London.

IV.B.7 Sodium and Hypertension

Historically, dietary salt (sodium chloride) has been obtained by numerous methods, including solar evaporation of seawater, the boiling down of water from brine springs, and the mining of “rock” salt (Brisay and Evans 1975). In fact, R. P. Multhauf (1978) has pointed out that “salt-making” in history could be regarded as a quasi-agricultural occupation, as seen in frequent references to the annual production as a “harvest.”

Such an occupation was seasonal, beginning with the advent of warm weather or the spring high tide and ceasing with the onset of autumnal rains. Multhauf has argued further that the quest for salt led to the development of major trade routes in the ancient world. The historian Herodotus, for example, described caravans heading for the salt oases of Libya, and great caravan routes also stretched across the Sahara, as salt from the desert was an important commodity exchanged for West African gold and slaves. Similarly huge salt deposits were mined in northern India before the time of Alexander the Great, and in the pre-Columbian Americas, the Maya and Aztecs traded salt that was employed in food, in medicines, and as an accessory in religious rituals. In China, evidence of salt mining dates from as early as 2000 B.C.

Homer termed salt “divine,” and Plato referred to it as “a substance dear to the gods.” Aristotle wrote that many regarded a brine or salt spring as a gift from the gods. In the Bible (Num. 18:19), it is written: “This is a perpetual covenant of salt before the Lord with you and your descendants also.” In the Orient, salt was regarded as the symbol of a bond between parties eating together. In Iran, “unfaithful to salt” referred to ungrateful or disloyal individuals. The English word “salary” is derived from *salarium*, the Latin word for salt, which was the pay of Roman soldiers. Moreover, Roman sausages were called *salsus* because so much salt was used to make them (Abrams 1983).

The preservative properties of salt have maintained the essentiality of the mineral throughout history. It helped meat last over long journeys, including those of marching armies and the migrations of peoples. Salt's power to preserve meat led to the so-called invention of salted herring in fourteenth-century Europe, which has been called “a new era in the history of European salt production” (Multhauf 1978: 9). The technique of pickling preserves food by extracting water from animal tissues, making the dehydrated meat or fish resistant to bacterial attack (Bloch 1976).

During the eighteenth century, other industrial uses began to be found for salt. The invention in 1792 of a way to make sodium carbonate began the carbonated-water industry, and by 1850, 15 percent of the salt of France was going into soda. Since that time, nondietary uses of salt have far outweighed its employment for culinary purposes (Multhauf 1978).

Historically, governments appreciated the importance of salt and have taxed it since ancient times (Multhauf 1978). During the nineteenth century in the United States, a salt tax helped build the Erie Canal, and during the twentieth century in India, Mahatma Gandhi revolted against a salt tax, leading to the famous “March to the Sea.” Such has been the importance of salt that one historian has written: “Clearly, anyone who can control the salt supply of a community has powers of life and death. The control of water, being more ubiquitous than salt, is not so simple to put into effect” (Bloch 1976: 337).

Salt and Sodium: Essential to Life

Salt in the Body

In 1684, Robert Boyle became the first to demonstrate scientifically that the “salty taste” in blood, sweat, and tears was actually caused by the presence of salt. After he removed the organic matter from whole blood by ignition, a fixed salt remained, which he found to be virtually identical to marine salt. About a century later (1776), H. M. Rouelle showed that a large proportion of the inorganic materials in blood serum could be isolated in the form of cubic crystals of “sea salt.” Later still, in the nineteenth century, J. J. Berzelius and A. J. G. Marcet revealed that sodium chloride was the principal inorganic constituent of other body fluids (both those that occasionally collected in the abdominal cavity, around the lungs or heart, or in a cyst or a blister and those that permanently surrounded the brain and spinal cord) and was present in much the same concentration as in blood serum (Kaufman 1980). In the same century (1807), Sir Humphry Davy discovered both sodium and potassium by passing an electrical current through moist caustic potash or caustic soda. More recently, biomedical researchers have defined sodium as the principal cation of the circulating blood and tissue fluids of animals (Denton 1982).

Sodium is the sixth most common element on earth. Sodium chloride (what we commonly call salt) is the chemical combination of ions of sodium (Na^+ , molecular weight 23) and chlorine (Cl^- , molecular weight 35.5) – the latter element, in its pure form, is a deadly greenish-yellow gas that reacts with water to form hydrochloric acid. Forty percent of the weight of common salt is made up of sodium; the remainder is chloride. Pure sodium is never found in nature. When freed from common table salt by electrolysis, sodium is a soft metal, lighter than water, and so reactive with oxygen in the air that it must be specially stored in air-free containers to prevent it from exploding. Sodium also reacts violently with water, as the two together form sodium hydroxide, in the process liberating hydrogen gas, which, in turn, bursts into flame from the heat of the reaction.

Nonetheless, even though a reactive element, sodium is essential to animal and human life. Indeed, life could be defined as the sum of the chemical processes that take place in the solution of salts between and within cells. In humans, the nutrients required to fuel life processes are first chewed and mixed with salt solutions produced by the salivary glands, then dissolved in salt and enzyme solutions from the stomach and pancreas, absorbed as salt solutions from the intestines, and delivered to the cells dissolved in a salt solution that ultimately depends on the ingestion of critical amounts of sodium and water. Excreted body fluids – blood, sweat, and tears – and feces are made up of these salts, and sodium salts are their key ingredient.

The Physical Need for Salt

In the nineteenth century, G. Bunge made the observation that carnivores never sought out salt, but herbivores did. His observation seemed to fit into common knowledge – hunters as well as husbandry-men knew that herbivores came to salt licks – but Bunge suggested something new: that salt was a necessity for life. In his travels he observed in numerous places that carnivores never ate salt but herbivores seemed to have a vital need for a supplement of it. He noted that herbivores excreted 3 to 4 times as much potassium as carnivores and theorized that the much higher potassium content of the vegetarian diet displaced sodium from body salts, causing an increase in the amount excreted in the urine. Therefore he reasoned that continuous consumption of a purely vegetarian diet with large amounts of potassium would make a large intake of sodium necessary for the maintenance of sodium balance (Bunge 1902).

Decades later, anthropologist Alfred Kroeber took issue with this notion of a biologically driven hunger for salt. He observed the Native Americans living along the Pacific coast of the United States and noted that salt was consumed in the south but not in the north. He saw no relationship among such factors as dietary salt use, the relative prevalence of seafood or meat in the diets, and the various climatic conditions, writing: “It must be concluded that whatever underlying urge there may be in physiology as influenced by diet and climate, the specific determinant of salt use or nonuse in most instances is social custom, in other words, culture” (Kroeber 1942: 2). H. Kaunitz found similar situations in areas of Australia, South Africa, and South America and suggested that salt craving might arise from “emotional” rather than innate needs (Kaunitz 1956).

J. Schulkin (1991) has recently noted that among psychologists in the 1930s, it was widely believed that “learning” – not biologically driven physical need – was responsible for the ingestion of minerals. However, C. P. Richter (following American physiologist Walter Cannon’s theory of *The Wisdom of the Body* [1932]) took a minority view that “learning” might not be the primary driver, and in 1936, he provided the first experimental evidence for a salt appetite. He removed the adrenal glands from experimental rats, thus depriving them of the sodium-retaining hormone aldosterone – a situation that would prove fatal in the absence of dietary sodium – and the amount of sodium ingested by the adrenalectomized rats increased dramatically. In 1939, he hypothesized that the drive for sodium was innate, and in 1941, he “discovered” that hormonal signals generate sodium hunger. Moreover, in 1956, he showed that during reproduction periods, the ingestion of salt by female rats rose markedly (Schulkin 1991).

One of the key experimental laboratories in the study of sodium metabolism has been the Howard

Florey Institute in Melbourne, Australia, where investigators, under the leadership of Derek Denton (1982), have conducted many experiments. In one of the most notable, researchers trained sheep to press a lever to get salt water to drink. The sheep were then depleted of salt by saliva drainage. When finally given access to the lever for salt water, the sheep within 30 minutes consumed the precise amount of sodium that they had lost. Their “wisdom of the body” was such that, even if the sheep were given salt solutions of varying concentrations, they still consumed the amount required to replace the deficit.

The work of Denton and others strongly supported the view that there is an innate hunger for salt and that the brain controls this behavior (Schulkin 1991). But although there are a number of minerals that are “essential” nutrients, only sodium seems to command a “built-in” hunger; there is no “innate” craving for magnesium or potassium, to choose two examples. On the other hand, it is likely that, in the past, the hunger for sodium abetted the intake of other essential minerals, which would usually have been found in the same “salt-licks” as sodium. Sodium hunger does not *require* “learning,” although significant “learning” does interact with innate mechanisms to help guide sodium-hunger behavior (Schulkin 1991). The following is a summary of Schulkin’s ideas of the steps involved in the innate sodium-hunger pathway:

1. An animal is “sodium-hungry.” (This would result from either a reduction in salt intake or excessive excretion of sodium from nonrenal sources, such as intestines or sweat glands.)
2. A “representation” of salty taste is activated in the brain.
3. The representation serves to guide the animal’s behavior in its search for salt – including its location, identification, and ingestion of the mineral.
4. Innate mechanisms are responsible for the sodium-hungry animal:
 - a. Ingesting the salt immediately upon its first exposure (no “learning” is required for this), and
 - b. Noting the significance of salt when not sodium-hungry.
5. Thus, in terms of b, there is a hedonic shift in the perception of salt that emerges in the salt-hungry animal.
6. The result is a motivated behavior with appetite (physiologic need) and consummatory phases (behavioral want) in search of salt.

Sodium Physiology

Sodium is vital in maintaining the pressure and volume of the blood and the extracellular fluid. A major purpose of the blood is to carry intracellular fluids, bringing nutrients to the cells and removing metabolic products from them. As blood flows through the capillaries, water – containing nutrients – passes from the capillaries into the extracellular spaces to

bathe the cells with nutrients and pick up cellular metabolic products (mostly waste), which are then swept by water movement back into the veins and carried to the kidney, liver, and lungs for metabolism or excretion. Sodium is also important in the transmission of nerve impulses, helps to metabolize carbohydrates and proteins, and has independent interactions with other ions – such as those of potassium, calcium, and chlorine – to maintain the “sea within” us. But most importantly, from a medical viewpoint, sodium is a vital factor in the regulation of blood pressure.

Sodium is measured in units of moles or grams. For nutritional purposes, grams are used, usually milligrams (1 gram [g] = 1,000 milligrams [mg]); for clinical purposes (to measure concentration), millimoles per liter are used (1 mole = 1,000 millimoles or mmols). One mole equals the molecular weight of the element. As the atomic weight of sodium is 23, 1 mole of sodium is equal to 23 grams of sodium (23,000 mg), and 2,300 mg of sodium is the same as 100 mmols of sodium.

Until the late 1940s, the measurement of sodium in both biological fluids and diets was a mostly laborious process that required the skills of a quantitative analytical chemist using 13 different steps – including, among others, “ashing,” extracting, evaporating, precipitating, washing, and weighing sodium yields on a microbalance – to determine the quantity in a single sample (Butler and Tuthill 1931). But in 1945, a revolution in the analytic accuracy and speed of sodium measurement was begun with the first report of the use of the flame photometric method. In 1945, this technique still required precipitation of plasma before analysis, but by 1947, only dilution of plasma was required (Overman and Davis 1947), and by 1949, instruments were available that could provide very accurate results within 5 minutes using either plasma or urine. By 1953, these devices were in wide use (Barnes et al. 1945; Mosher et al. 1949; Wallace et al. 1951).

The body has built-in “set points” designed to maintain sodium in homeostasis. When it takes in less salt than is lost in the urine, sweat, and stool, the concentration of sodium in the blood falls. When the blood sodium falls below an inherited “set point” (about 140 mmol per liter of serum), an area of the brain that is bathed by blood senses the decreased sodium concentration and activates hormonal defenses to maintain a constant concentration of the mineral. If the concentration of sodium continues to diminish, the kidneys will adjust by accelerating the excretion of water, so that the blood’s sodium concentration is maintained at the vital level. If the sodium supply is not replenished, there is a gradual desiccation of the body and, finally, death. In other words, a lack of sufficient sodium causes the organism literally to die of thirst.

By contrast, if blood sodium increases above the set-point level, a secretion of antidiuretic hormone (ADH) is released by the pituitary gland, and thirst

mechanisms are activated to find and ingest water until the sodium concentration is reduced. At the same time, ADH causes the kidneys to excrete less water in an attempt to keep the body's sodium at the correct concentration. If, however, the water supply is not replenished, more sodium will be excreted, and eventually, these water losses will lead to death.

The overriding mechanism that regulates total body sodium (and blood pressure) has been termed the "renal-fluid volume mechanism for pressure control" by A. C. Guyton and colleagues (1995). An analysis of the factors controlling blood pressure has shown that it can only be raised by one of two mechanisms: increasing the intake of dietary salt or limiting the kidney's ability to excrete sodium.

Sodium and Human Evolution

The body's need for sodium may also have played a role in genetic variability within the human species. During the 1980s, theories were proposed that suggested such a role in two diseases related to salt metabolism: cystic fibrosis and hypertension.

Cystic fibrosis (CF) is a recessive genetic condition related to sodium metabolism, in which, it was hypothesized, the carrier state had been protective of fluid and electrolyte loss during epidemics of diarrhea in human history. CF carriers, notably children before the age of reproduction, were thought to have protective mechanisms that diminished the loss of water during episodes of infectious diarrhea. Thus, individuals who were genetically enabled to control water and salt losses were more likely to survive to reproductive age. Indeed, the heterozygote carrier has been shown to have less sodium loss in feces than the homozygote noncarrier (Gabrial et al. 1994).

A more controversial evolutionary hypothesis is that one form of hypertension (high blood pressure) – "salt-sensitive" hypertension, which has a high frequency among African-Americans – may result, in part, from genetic adaptation to the African environment and its diseases of the past. More specifically, it has been suggested that – both during the trans-Atlantic slave trade and during the period of slavery itself – individuals able to conserve sodium would have been more likely to survive the dehydrating diseases aboard ship, as well as the debilitation of hard physical labor. If so, then this past experience might be partially responsible for today's prevalence of "salt-sensitive" high blood pressure among black people in the Western Hemisphere (Wilson and Grim 1991; Curtin 1992; Grim and Wilson 1993).

Salt Deficiency

When humans go without salt in the diet, or lose it because of illness, the major symptoms are apathy, weakness, fainting, anorexia, low blood pressure, and, finally, circulatory collapse, shock, and death. Sir William Osler (1978: 121–2), observing dehydrated

cholera patients in the late nineteenth century, provided a classic description of the condition:

[P]rofuse liquid evacuations succeed each other rapidly . . . there is a sense of exhaustion and collapse . . . thirst becomes extreme, the tongue white: cramps of great severity occur in the legs and feet. Within a few hours vomiting sets in and becomes incessant. The patient rapidly sinks into a condition of collapse, the features are shrunk, the skin of an ashy gray hue, the eye-balls sink in the sockets, the nose is pinched, the cheeks are hollow, the voice becomes husky, the extremities are cyanosed, and the skin is shriveled, wrinkled and covered with a clammy perspiration. . . . The pulse becomes extremely feeble and flickering, and the patient gradually passes into a condition of coma.

Many cholera patients in the past could have been saved with rehydration therapy, and it is a central tenet in modern medical treatment that lost body fluids should be replaced with others of the same composition. Replacing a salt loss by giving water or a water loss by giving salt can be fatal. Although a history of the illness and an examination of the patient can provide clues to the type of loss, the best method is to test the blood and urine chemically – a method that only became possible in the 1930s, with the most useful test that which determined the amount of chloride in urine. Accomplished by simply mixing 10 drops of urine with one drop of an indicator and then adding silver nitrate, a drop at a time, until the end point was reached, this was called the "Fantus test" after Dr. Bernard Fantus at the University of Chicago.

This test proved so useful in treating salt- and water-depleted British soldiers in India and Southeast Asia during the mid-1930s that Dr. H. L. Marriott, in his classic text on fluid replacement therapy, stated: "It is my belief that the means of performing this simple test should be available in all ward test rooms and in every doctor's bag" (Marriott 1950: 56).

Most early studies of sodium depletion in humans were prompted by diseases. One was Addison's disease (in which the adrenal gland that makes the sodium-retaining hormone for the body stops working), and another was diabetes (in which a high level of blood glucose forces the excretion of large amounts of water by the kidneys). Such studies were also conducted in cases of extreme depletion brought on by starvation or acute diarrhea. In the 1930s, however, R.A. McCance (1935–6) published his report of a series of experiments that established a baseline on the clinical nature and physiology of sodium depletion in humans.

To induce salt depletion, McCance (1935–6) employed a sodium-free diet combined with sweating. (Because laboratory animals do not sweat, he

used humans as his test subjects.) There was no “research-quality” kitchen available, the food was prepared in the McCance home, and the subjects of the experiment – all volunteers – slept and ate there. The diet consisted of sodium-free “casein” bread, synthetic salt-free milk, sodium-free butter, thrice-boiled vegetables, jam, fruit, homemade sodium-free shortbread, and coffee. During recovery periods, the volunteers ate weighed quantities of high-sodium foods (such as anchovies and bacon) and small, weighed amounts of sodium chloride. Sweating was induced by placing the subjects in a full-length radiant heat bath – for two hours with the heat on and then 10 minutes with the heat off. Their sweat was collected in rubber sheets, and a final washing of each subject with distilled water ensured that even small amounts of lost sodium would be accounted for. The subjects’ average sweat loss was 2 liters, and they commented that the washing procedure was “not uncomfortable” after 2 hours in the hot bath (McCance 1935–6).

By reducing sodium in the diet, along with inducing sodium losses through sweating, McCance and his colleagues found that only a week was required to make healthy subjects seriously sodium depleted. They maintained 4 volunteers in this condition for an additional 3 to 4 days, so that the total period of deprivation lasted about 11 days.

Detailed measurements of intake (food and water) and output (sweat, urine, and feces) recorded that the subjects lost 22.5 g of sodium and 27.2 g of chloride – or about 50 g of salt. Their body weights dropped by about 1 kilogram (kg) per day, and sodium excretion averaged 3,400 mg of sodium per day for the first 4 days. Weights then stabilized, but sodium loss continued.

As the deficiency progressed, the volunteers all experienced feelings of physical fatigue, anorexia, nausea, difficulty in urinating, and extremely weak pulses. Muscle spasms and cramps – especially cramps in the fingers – were common. The subjects’ faces became drawn and “ill-looking,” and they slowed mentally, becoming dull and apathetic. McCance was struck by the similarity of a number of these symptoms to those of Addison’s disease, but the symptoms and signs all rapidly cleared up when the volunteers resumed consumption of sodium (McCance 1935–6).

Both before and during World War II, as many in the Allied armed forces were severely disabled by heat- and water-related illnesses, there was intense interest in understanding the mechanics of water and salt metabolism and the effects of heat on the human body. Research was even undertaken to see how long a man could survive on a raft in the ocean, or in the desert, so that restrictions could be placed on certain military activities (such as limiting the duration of searches for lost aviators, who, after the specified survival time had passed, might reasonably be presumed dead). Other studies examined the conditions that servicemen could be forced to work under and defined safe limits. For

example, after 5 hours of marching in the heat, with full packs and no water, even well-conditioned marines could not continue (Ladell 1949).

During and after World War II, there was also interest in the effects of diarrhea. J. L. Gamble (1945) showed that intestinal secretions contained more sodium than chloride, and D. A. K. Black (1946) reported a series of experiments on 10 men with blood pressures averaging 94 mm Hg SBP/59 mm Hg DBP, who were victims of tropical sprue (a disease characterized by chronic diarrhea). The patients were bedridden, listless, and incapable of exertion. But these symptoms disappeared – and blood pressure rose to normal – with sodium supplementation (Black 1946).

Excess Sodium

Humans, as noted, have evolved complex “redundant systems” to regulate sodium and other essential minerals. For marine animals, deriving sodium from the sea was a relatively easy matter. As evolution progressed, however, satisfaction of sodium needs became a more complicated task. Land dwellers had first to locate sources of sodium, then ingest the mineral, and, further, conserve it within their bodies. To achieve this, physiological and behavioral mechanisms evolved that were designed primarily to protect against a life-threatening deficit of sodium, such as can occur with vomiting, sweating, diarrhea, or kidney malfunction.

But although the body’s systems are reasonably effective against sodium deficit, evolution did not do as well in protecting humans against an excessive intake of the mineral. There are two different kinds of excessive sodium intake: (1) acute ingestion of salt without water, or of very salty water (such as seawater or other briny water); and (2) chronic ingestion of toxic levels of sodium in the food supply.

It seems likely that the former was never a major problem in the history of humankind and probably occurred only when people felt forced to drink seawater or the water from salt springs or salt lakes. Chronic ingestion of excess salt in food, however, is both a recent and a very real problem. Until the past few centuries, salt intake was primarily determined by the amount a person chose to add to food (“active intake”). Increasingly, however, as foods were preserved in salt, and especially today with foods processed with salt, the great majority of salt intake has become “passive,” meaning that food processors and manufacturers – not consumers – decide the quantity of salt to be included in the foods they produce.

Indeed, it has been estimated that in prehistoric times, the daily human intake of sodium was about 690 mg, with 148 mg derived from vegetables and 542 mg from meat (Eaton and Konner 1985). By contrast, today the U.S. Food and Drug Administration (FDA) recommends keeping dietary intake to

2,400 mg of sodium per day (the amount contained in 6 g - or about 1 teaspoon - of table salt). This is roughly the same amount accepted by an advocacy group that promotes a low-salt diet, Consensus Action on Salt and Hypertension (CASH). Needless to say, even such a "low-salt" diet still delivers 3.5 times the amount of sodium provided by the meat and vegetables of the Paleolithic diet.

Although there is also concern over high salt intake by children, food processors only relatively recently halted their practice of adding sodium to baby food. (Presumably, the mineral was meant to improve the food's flavor for parents.) This change for the better followed the observation by L. K. Dahl, M. Heine, G. Leitl, and L. Tassinari (1970) that young rats with a high sodium intake became hypertensive and remained so for the rest of their lives. But despite this discovery, the average sodium intake by two-year-olds in the United States remains higher than the amount recommended by the FDA and by CASH (Berenson et al. 1981).

Moreover, the average intake throughout much of the industrialized world today is about 10 g of table salt (3,900 mg of sodium) per person per day (James, Ralph, and Sanchez-Castillo 1987; USDA/USDHHS 1995). But only about 10 percent of the sodium consumed occurs naturally in foods; another 15 percent is added by consumers ("active intake"), and the remaining 75 percent is added to food by manufacturers and processors. Therefore, based upon the average industrial diet, only 390 mg of sodium is naturally occurring, 585 mg is added by the consumer, and a substantial 2,925 mg is derived ("passive intake") from sodium added during processing (James et al. 1987).

Obviously, then, a low-salt diet like that of our ancient ancestors seems an impossible goal; increasingly, the population is at the mercy of the food industry. Yet, that industry gains several advantages by using excessive salt in food processing. Salt, added to food, improves flavor and palatability, especially for those who have become addicted to its taste. Salt increases the "shelf-life" of food (although, with today's effective packaging and refrigeration technology, there is little need for the presence of near-toxic levels of sodium in the food supply), and salt adds very inexpensive weight to the final product (by helping to retain water), thereby increasing profit margins. Finally, increased salt consumption stimulates thirst and thus increases consumption of the beverages sold by many of the major food companies (McGregor 1997).

History of Hypertension

"Blood pressure" refers to the force exerted against the walls of the arteries as blood is pumped throughout the body by the heart. This pressure can be measured, and abnormally high levels indicate a condition of "high blood pressure," or hypertension, which is

classified into two types. *Secondary* hypertension is that resulting from a known cause, such as a disease of the kidneys, whereas *essential* or *primary* hypertension arises with no evident cause. More than 90 percent of hypertension cases fall into the latter category (Wilson and Grim 1993).

Salt consumption was linked with blood pressure long ago - as early as the second century B.C. in ancient China. In the *Huang-ti nei-ching* ("The Inner Classic of the Yellow Emperor"), it was written that "if too much salt is used for food, the pulse hardens" (Swales 1975: 1).

In 1836, Richard Bright reported on the kidneys and hearts of 100 patients who had died of kidney problems. He noted that, in most instances, when the kidney was small and shrunken, the heart was often markedly enlarged - an indication of high blood pressure (Bright 1836). In the late nineteenth century, R. Tigerstedt and T. G. Bergman (1898) coined the term "renin" in their report that saline extracts from kidney tissue raised blood pressure. A few years later, French researchers L. Ambard and E. Beaujard noted that blood pressure was lowered by the dietary restriction of sodium chloride and raised by its addition or increase; they believed, however, that chloride was the culprit because they were unable to measure sodium content (Ambard and Beaujard 1904).

Many attempts to induce high blood pressure in experimental animals followed but proved inconclusive. In 1934, however, Harry Goldblatt noted that when he placed an adjustable clamp so that it partially blocked the renal artery of a dog, the animal developed a rapid increase in blood pressure that was sustained so long as the clamp remained in place. Research to determine the mechanism causing this high blood pressure began, and by 1940, two teams of investigators (Eduardo Braun-Menendez and colleagues in Buenos Aires and Irving Page and O. M. Helmer in Indianapolis) succeeded in isolating a material that caused vasoconstriction. Both groups reported that there was a substance - coming from the kidneys - that, when mixed with blood, generated a potent vasoconstricting and blood-pressure-raising substance. The South American group called this "hypertensin" and the U.S. group "angiotonin," but eventually the two were determined to be the same chemical, which was named "angiotensin."

Within a few years, W. Kempner (1944) reported that a low-sodium diet decreased blood pressure and heart size even in cases of malignant hypertension. The following year, he described the effects of a low-sodium diet on 65 patients who were hypertensive but showed no evidence of renal disease. After an average of only 48 days on a rice-fruit diet, the average blood-pressure readings decreased from a systolic blood pressure (SBP) of 197 mm Hg and a diastolic blood pressure (DBP) of 115 mm Hg to an SBP of 151 mm Hg and a DBP of 97 mm Hg. In those patients who experienced a decrease in blood pres-

sure, the response was obvious within the first 7 to 10 days on the low-sodium diet, and Kempner observed that the maximum decrease in blood pressure was first attained after only 10 days.

Most early blood-pressure studies were on whites, yet the prevalence of high blood pressure in blacks is much greater and, as mentioned, is thought to result, at least in part, from salt intake. The first report of an association between salt intake and blood pressure in African-Americans was by A. Grollman and colleagues (1945), who studied patients given less than 1 gram (<1,000 mg) of sodium chloride in their daily diets. In the case of two black women, blood pressure declined to normal, promptly rose when salt intake was increased, and then fell again when the low-sodium diet was resumed.

Five years later, V. P. Dole and colleagues (1950) reported the results of a series of studies to evaluate the sodium content of Kempner's rice-fruit diet and its effect on blood pressure and heart size. They confirmed all of Kempner's observations and, further, documented that the effect of the diet was related to its low sodium content and not low chloride (Dole et al. 1950).

The introduction of diuretics, which act to excrete sodium and water via the kidneys, came in the late 1950s, and these were quickly shown to lower blood pressure significantly (Freis et al. 1958). In the 1970s, however, a series of observations made by a research group in Indianapolis (which included Clarence E. Grim, co-author of this chapter) once again focused attention on the relationship between dietary sodium and blood pressure. This culminated in several reports indicating that even a normotensive individual would experience an increase in blood pressure if enough salt was consumed (Murray et al. 1978; Luft et al. 1979). In addition, ethnic differences in sodium metabolism and blood-pressure responses were documented in normotensive subjects; such evidence demonstrated, for the first time, that blacks were more sensitive to salt than whites. These studies also demonstrated the enormous capacity of human kidneys to excrete sodium, which was proof of Guyton's hypothesis that excess salt will always increase blood pressure (Guyton et al. 1995).

Current Thinking on Sodium and Hypertension

Cardiovascular disease is now a major cause of death in all countries of the world, and high blood pressure is the most common (although treatable) precursor of heart disease. Unfortunately, as pointed out in a recent review by J. P. Midgley, A. G. Matthew, C. M. T. Greenwood, and A. G. Logan (1996), most sodium-reduction trials in hypertensive subjects have not yet produced definitive evidence that reducing sodium intake improves long-term health (but see also Staessen et al. 1997). Indeed, some have argued

(based on a single observational study) that reducing salt intake may do more harm than good (Alderman et al. 1995).

Nonetheless, high dietary salt intake has been reported to be associated with other adverse medical outcomes, including death from stroke, enlargement of the heart (a precursor to congestive heart failure), and even asthma mortality (Antonios and McGregor 1995). Moreover, recent population-based studies seem to confirm that the major cause of high blood pressure is an excessive dietary intake of sodium. Using a standardized protocol in a 52-center worldwide study, INTERSALT showed a positive, significant, linear relationship between salt intake and blood pressure. The statistical relationship suggested that each 100 mmol of sodium intake was responsible for a 2 mm Hg increase in systolic blood pressure. In addition, the INTERSALT findings suggest a powerful effect of dietary sodium intake on the rise in blood pressure with age. It has also been argued that a reduction in the average intake of sodium in Britain – from 150 mmol to 100 mmol per day – could reduce strokes and heart attacks in that nation by 22 percent and 16 percent, respectively, and would have a greater impact than that of all of the drugs used to treat high blood pressure (McGregor and Sever 1996).

For the majority of hypertensive persons, it seems well established that lifestyle improvements, such as lowering dietary sodium intake while increasing dietary potassium intake, reducing body weight, and increasing exercise, can lower blood pressure. Although such efforts would likely lower a person's blood pressure by "only" 5 mm Hg, it is important to consider the societal health benefits of such a downward shift across the entire population. From a population-wide perspective, this could dramatically reduce the prevalence of hypertension and cardiovascular disease.

It has been estimated that it costs \$1,000 per year to treat each hypertensive person (Elliott 1996). If lifestyle changes could lower the average blood pressure by only 5 mm Hg, then 21.4 million persons would no longer require treatment. This would save the United States about \$21 billion a year in the cost of health care for hypertension alone, not to mention the costs saved by the reduction of strokes by about 40 percent and cardiovascular disease by about 25 percent. The potential economic advantage of implementing low-cost lifestyle changes to lower blood pressure across society should be obvious. Indeed, it is clear that most countries of the world will not be able to afford expensive medical therapies to control high blood pressure and should be implementing public-health strategies to lower blood pressure and the devastating consequences of hypertension for a country's workforce.

Thomas W. Wilson
Clarence E. Grim

Bibliography

- Abrams, H. L. 1983. Salt and sodium: An anthropological cross-cultural perspective in health and disease. *Journal of Applied Nutrition* 35: 127-58.
- Alderman, M. H., S. Madhavan, H. Cohen, et al. 1995. Low urinary sodium is associated with greater risk of myocardial infarction among treated hypertensive men. *Hypertension* 25: 1144-52.
- Ambard, L., and E. Beaujard. 1904. Causes de l'hypertension artérielle. *Archives Générales de Médecine* 1: 520-3.
- Antonios, T. F., and G. A. McGregor. 1995. Deleterious effects of salt intake other than effects on blood pressure. *Clinical and Experimental Pharmacology and Physiology* 22: 180-4.
- Barnes, R. B., D. Richardson, J. W. Berry, and R. L. Hood. 1945. Flame photometry: A rapid analytical procedure. *Industrial Engineering Chemistry, Analytical Edition U.S.* 17: 605-11.
- Berenson, G. S., S. R. Srinivasan, G. C. Frank, and L. S. Webber. 1981. Serum lipid and lipoprotein in infants and children and their relationship with diet. *Progress in Clinical & Biological Research* 61: 73-94.
- Black, D. A. K. 1946. Salt deficiency in sprue. *Lancet* 2: 671-5.
- Bloch, M. R. 1976. Salt in human history. *Interdisciplinary Science Reviews* 1: 336-52.
- Bright, R. 1836. Tabular view of the morbid appearances in 100 cases connected with albuminous urine, with observations. *Guy's Hospital Reports* 1: 380-400.
- Brisay, K. W. de, and K. A. Evans, eds. 1975. *Salt: The study of an ancient industry*. Colchester, England.
- Bunge, G. von 1902. *Text-book of physiological and pathological chemistry*. Second English edition. Translated from the fourth German edition by Florence A. Starling and edited by Ernest H. Starling. Philadelphia, Pa.
- Butler, A. M., and E. Tuthill. 1931. An application of the uranyl zinc acetate method for determination of sodium in biological fluids. *Journal of Biological Chemistry* 93: 171-80.
- Cannon, Walter Bradford. 1932. *The wisdom of the body*. New York.
- Curtin, P. D. 1992. The slavery hypothesis for hypertension among African-Americans: The historical evidence. *American Journal of Public Health* 82: 1681-6.
- Dahl, L. K., M. Heine, G. Leidl, and L. Tassinari. 1970. Hypertension and death from consumption of processed baby foods by rats. *Proceedings of the Society of Experimental Biological Medicine* 133: 1405-8.
- Denton, D. 1982. *The hunger for salt: An anthropological, physiological, and medical analysis*. Berlin.
- Dole, V. P., L. K. Dahl, G. C. Cotzias, et al. 1950. Dietary treatment of hypertension. Clinical and metabolic studies of patients on the rice-fruit diet. *Journal of Clinical Investigation* 29: 1189-1206.
- Eaton, B. S., and M. Konner. 1985. Paleolithic nutrition: A consideration of its nature and current implications. *New England Journal of Medicine* 312: 283-9.
- Elliott, W. J. 1996. The costs of treating hypertension: What are the long-term realities and pharmacoeconomics? *Postgraduate Medicine* 99: 241-8.
- Freis, E. D., A. Wanko, I. M. Wilson, and A. E. Parrish. 1958. Chlorothiazide in hypertensive and normotensive patients. *Annals of the New York Academy of Science* 71: 450.
- Gabrial, S. E., K. N. Brigman, B. H. Koller, et al. 1994. Cystic fibrosis heterozygote resistance to cholera toxin in the cystic fibrosis mouse model. *Science* 266: 107-9.
- Gamble, J. L. 1945. Congenital alkalosis. *Journal of Pediatrics* 26: 509-18.
- Grim, C. E., and T. W. Wilson. 1993. Salt, slavery, and survival: Physiological principles underlying the evolutionary hypothesis of salt-sensitive hypertension in Western Hemisphere blacks. In *The pathophysiology of hypertension in blacks*, ed. J. C. S. Fray and J. G. Douglas, 22-44. New York.
- Grollman, A., T. R. Harrison, M. F. Mason, et al. 1945. Sodium restriction in diet for hypertension. *Journal of the American Medical Association* 129: 533-7.
- Guyton, A. C., J. E. Hall, T. G. Coleman, et al. 1995. The dominant role of the kidneys in long-term arterial pressure regulation in normal and hypertensive states. In *Hypertension: Pathophysiology, diagnosis, and management*. Second edition, ed. J. H. Laragh and B. M. Brenner, 1311-26. New York.
- James, W. P. T., A. Ralph, and C. P. Sanchez-Castillo. 1987. The dominance of salt in manufactured food in the sodium intake of affluent societies. *Lancet* 1: 426-9.
- Kaufman, Dale W., ed. 1980. *Sodium chloride: The production and properties of salt and brine*. Washington, D.C.
- Kaunitz, H. 1956. Causes and consequences of salt consumption. *Nature* 178: 1141-4.
- Kempner, W. 1944. Treatment of kidney disease and hypertensive vascular disease with rice diet. *North Carolina Medical Journal* 5: 125-33.
- Korotkoff, N. S. 1905. On methods of studying blood pressure. *IZV Voenno-medakad* 11: 365.
- Kroeber, A. L. 1942. Culture element distribution: XV. Salt, dogs, tobacco. *Anthropological Records* 6: 1-20.
- Ladell, W. S. S. 1949. The changes in water and chloride distribution during heavy sweating. *Journal of Physiology* 108: 440-50.
- Luft, F. C., L. I. Rankin, R. Bloch, et al. 1979. Cardiovascular and humoral responses to extremes of sodium intake in normal white and black men. *Circulation* 60: 697-706.
- Marriott, H. L. 1950. *Water and salt depletion*. Springfield, Ill.
- McCance, R. A. 1935-6. Experimental sodium chloride deficiency in man. *Proceedings of the Royal Society of London* 119: 245-68.
- McGregor, G. A. 1997. President's symposium: Salt, blood pressure, and human health. *American Journal of Hypertension* 10: 220A.
- McGregor, G. A., and P. S. Sever. 1996. Salt - overwhelming evidence but still no action: Can a consensus be reached with the food industry? *British Medical Journal* 312: 1287-9.
- Midgley, J. P., A. G. Matthew, C. M. T. Greenwood, and A. G. Logan. 1996. Effect of reduced dietary sodium on blood pressure: A meta-analysis of randomized controlled trials. *Journal of the American Medical Association* 275: 1590-7.
- Mosher, R. E., A. M. Boyle, E. J. Bird, et al. 1949. The use of flame photometry for the quantitative determination of sodium and potassium in plasma and urine. *American Journal of Clinical Pathology* 49: 461-70.
- Multhauf, R. P. 1978. *Neptune's gift: A history of common salt*. Baltimore, Md.
- Murray, R., F. C. Luft, R. Bloch, and A. E. Weymen. 1978. Blood pressure response to extremes in sodium intake in normal man. *Proceedings of the Society of Experimental Biology* 159: 432-6.
- Osler, W. [1892] 1978. *The principles and practice of medicine*. Birmingham, England.
- Overman, R. R., and A. K. Davis. 1947. The application of flame photometry to sodium and potassium determinations

- in biological fluids. *Journal of Biological Chemistry* 168: 641–9.
- Schulkin, J. 1991. *Sodium hunger: The search for a salty taste*. Cambridge.
- Staessen, J. A., P. Lijnen, L. Theis, and R. Fagard. 1997. Salt and blood pressure in community-based intervention trials. *American Journal of Clinical Nutrition* 65 (Supplement): 661S–70S.
- Swales, J. D. 1975. *Sodium metabolism in disease*. London.
- Tigerstedt, R., and T. G. Bergman. 1898. Niere und Kreislauf. *Scandinavian Archives of Physiology* 7–8: 223–71.
- USDA/USDHHS (U.S. Department of Agriculture/U.S. Department of Health and Human Services). 1995. *Report of the Dietary Guidelines Advisory Committee on dietary guidelines for Americans*. Washington, D.C.
- Wallace, W. M., M. Holliday, M. Cushman, and J. R. Elkinton. 1951. The application of the internal standard flame photometer to the analysis of biologic material. *Journal of Laboratory and Clinical Medicine* 37: 621–9.
- Wilson, T. W., and C. E. Grim. 1991. Biohistory of slavery and blood pressure differences in blacks today: A hypothesis. *Hypertension* 17: 1122–8.
1993. Hypertension. In *The Cambridge world history of human disease*, ed. K. F. Kiple, 789–94. Cambridge and New York.

IV.B.8. 🦋 Other Trace Elements

Beginning early in the twentieth century, scientists were able to qualitatively detect small amounts of several elements in living organisms. In reports, these elements were often described as being present in “traces” or “trace amounts,” and within a short time, they became known as the trace elements. Today, trace elements are understood to be those elements of the periodic table that occur in the body in micrograms per gram ($\mu\text{g/g}$) of body weight or less. They may be essential – that is, they may be indispensable for growth, health, and completion of the life cycle – or they may be nonessential: fortuitous reminders of our geochemical origins or indicators of environmental exposure. Some of the nonessential trace elements can be beneficial to health through pharmacologic action, but all are toxic when consumed in excess.

An essential element is defined by many in the scientific community as an element whose dietary deficiency consistently results in a suboptimal function that is preventable or reversible by physiological amounts of the element. However, other experts accept an element as essential only if it has a defined biochemical function. Thus, there is no universally accepted list of trace elements that are considered essential.

The essential trace elements are usually required by humans in amounts indicated by milligrams per day (mg/d). In 1980, the term “ultratrace element” began to appear in the literature; this was defined as an element with a daily dietary requirement of some 50 nanograms per gram (ng/g) for animals (Nielsen 1980). For

humans, however, the term often is used to indicate elements with an established, estimated, or suspected requirement of less than 1.0 mg/d (generally indicated as micrograms per day, or $\mu\text{g/d}$) (Nielsen 1994b).

At present, both signs of deficiency and defined biochemical functions in humans are known for only 7 trace and ultratrace elements. The trace elements are copper, iron, and zinc, and the ultratrace elements are cobalt (in the form of cobalamin), iodine, molybdenum, and selenium. Although essential functions have been identified for manganese, unequivocal signs of its deficiency in humans have not been described. Conversely, such signs have been described for chromium and boron, but their specific biochemical functions have not been conclusively defined. Based on animal experiments, and discovery of their essential functions in lower forms of life, 4 other trace or ultratrace elements are probably essential to humans: arsenic, nickel, silicon, and vanadium. In addition, 2 elements with beneficial pharmacological properties are fluorine (anticariogenic) and lithium (antimanic). Since 1970, reports have appeared suggesting that other elements, including aluminum, bromine, cadmium, germanium, lead, rubidium, strontium, and tin, may be essential. However, weak, limited, or controversial data have prevented general acceptance of the essentiality of these elements.

The essential trace elements iron and zinc and the essential ultratrace element iodine are discussed in separate chapters and thus are not addressed here.

Arsenic

History of Oral Intake

Since ancient times, arsenicals have been characterized by actions both benevolent and malevolent. Very early, it was found that some arsenic compounds were convenient scentless and tasteless instruments for homicidal purposes, and until the nineteenth century, arsenic reigned as the king of poisons; indeed even today, the term “arsenic” is primarily thought of as synonymous with poison. Nonetheless, this bad reputation did not prevent it from becoming an important pharmaceutical agent. According to a review by H. A. Schroeder and J. J. Balassa (1966), the medicinal virtues of arsenicals were acclaimed by Hippocrates (460 to 357 B.C.), Aristotle (384 to 322 B.C.), Theophrastus (370 to 288 B.C.), and Pliny the Elder (A.D. 23 to 70).

Beginning in 1905, various arsenicals were developed for pharmacological purposes, and by 1937, they numbered 8,000. Arsenicals were considered at various times to be specific remedies in the treatment of anorexia, other nutritional disturbances, neuralgia, rheumatism, asthma, chorea, malaria, tuberculosis, diabetes, skin diseases, and numerous hematologic abnormalities. Today, however, either the use of arsenic for these disorders has fallen into disrepute or more effective alternatives have taken its place.

Discovery of Nutritional Importance

Reports describing attempts to produce a nutritional arsenic deficiency first appeared in the 1930s. E. Hove, C. A. Elvehjem, and E. B. Hart (1938) found that a milk diet fortified with iron, copper, and manganese, furnishing as little arsenic as 2 µg/day, did not significantly affect growth, hemoglobin concentration, and red blood cell number or fragility in rats. However, an arsenic supplement of 5 µg/day delayed the rate of fall of hemoglobin concentration in rats fed whole milk without mineral supplementation. J. T. Skinner and J. S. McHargue (1946) also found that arsenic supplementation increased hemoglobin concentrations in rats fed a skim milk and sucrose diet adequately supplemented with iron and copper.

H. A. Schroeder and J. J. Balassa (1966) found that rats and mice grew and developed normally when fed diets containing as little arsenic as 0.053 µg/g; additional arsenic in their drinking water improved the appearance of their skin and hair. Perhaps the first findings accepted as evidence for the possible nutritional essentiality of arsenic appeared in 1975 and 1976, when apparent deficiency signs were described for rats, goats, and pigs (Nielsen and Uthus 1984; Anke 1986; Uthus 1994). Subsequently, deficiency signs were reported for chickens and hamsters (Uthus 1994).

Dietary Importance

In goats, pigs, and rats, the most consistent signs of arsenic deprivation were depressed growth and abnormal reproduction characterized by impaired fertility and elevated perinatal mortality (Nielsen and Uthus 1984; Anke 1986). Arsenic deprivation also caused myocardial damage and death during lactation in goats. Some biochemical changes found with arsenic deprivation include depressed plasma triglycerides and taurine concentrations, liver polyamine concentration, liver glutathione S-transferase activity, liver S-adenosylmethionine concentration, and specific activity of S-adenosylmethionine decarboxylase (Uthus 1994). The large number of responses to arsenic deprivation – reported for a variety of animal species by more than one research group – strongly suggests that arsenic is a nutritionally important ultra-trace element for higher animals.

Although a biochemical function has not been defined for arsenic, its actions *in vitro* and the findings from animal deprivation studies indicate a function that affects the formation and utilization of labile methyl groups arising from methionine. Through this effect on methyl group metabolism, arsenic possibly affects the methylation of important molecules, such as DNA. If so, suboptimal methyl metabolism caused by lack of arsenic could be of concern; for example, hypomethylation of DNA can lead to some forms of cancer. It has also been suggested that arsenic plays an important role in human health because of findings that injuries of the central nervous system, vascu-

lar diseases, and cancers correlated with markedly depressed serum arsenic concentrations (Mayer et al. 1993). Because arsenic is most likely an essential nutrient, the belief that any form of it is unnecessary, toxic, or carcinogenic must be discarded.

Boron

History of Oral Intake

The recognition that the presence of boron in foods can be both beneficial and detrimental apparently dates from the 1870s. At that time, it was discovered that borax and boric acid could be used to preserve foods, and for the next 50 years or so, borate addition was considered one of the best methods of preserving or extending the palatability of foods such as fish, shellfish, meat, sausage, bacon, ham, cream, butter, and margarine (Gordon 1987). During World War I, the use of boron as a preservative played a vital role in preventing food crises. In other words, in the last decades of the nineteenth century and the first decades of the twentieth century, boron was considered beneficial, and prevailing medical opinion held that boron was innocuous because, despite its widespread use, no deaths had been blamed on it.

As early as 1902, however, some German and American scientists had begun to question this orthodox view. Foremost among the works that changed perceptions about boron was a report by H. W. Wiley (1904), which stated that consumption of boric acid in doses greater than 0.5 g/day for 50 days resulted in disturbances in appetite, digestion, and health in human volunteers. Following this change in perception, the opinion that boron posed a risk to health started to gain momentum, and by the mid-1920s, many countries began legislating against the addition of borates to food. Only during World War II were the new restrictions eased, as shortages made food preservation a major concern in many countries (Gordon 1987). After the war, restrictions were gradually reimposed and by the middle of the 1950s, the use of boron as a food preservative was virtually forbidden throughout the world. At this time, boron also began to receive attention from toxicologists. Because of some accidental poisonings and inappropriate uses in the medical profession, between 1950 and 1980 boric acid and borates were seen as only important to human health from a toxicological point of view.

Discovery of Nutritional Importance

During the 1920s, K. Warington (1923) and A. L. Sommer and C. B. Lipman (1926) showed that boron was an essential nutrient for plants, and in subsequent decades, other investigators attempted – unsuccessfully – to demonstrate that it was essential for higher animals as well (Hove, Elvehjem, and Hart 1939; Orent-Keiles 1941; Teresi et al. 1944). It was reported in 1945 that high dietary boron (100 to 1,000 µg/g) enhanced survival and increased body fat and liver

glycogen in potassium-deficient rats (Skinner and McHargue 1945), although further experiments, using a different diet with an unknown basal boron content and with different amounts of boron supplementation, failed to confirm these findings (Follis 1947). The inability in these early studies to produce a boron deficiency in animals apparently resulted in generations of students of biochemistry and nutrition learning that boron was a unique element because it was essential for plants but not for higher animals, including humans.

In the early 1980s, however, this dogma began to change. C. D. Hunt and E. H. Nielsen (1981) reported that boron deprivation depressed the growth of chicks, with the effect seemingly more marked when dietary cholecalciferol was marginally deficient. Moreover, morphological examination of the tibias of chicks indicated that an interaction between boron and cholecalciferol affected bone formation. In the years since this report, circumstantial evidence has accumulated to suggest strongly that boron is an essential nutrient for higher animals and humans.

Dietary Importance

Listing the signs of boron deficiency for animals is difficult because of the large number reported and because most studies of this deficiency in animals have used stressors to enhance the response to changes in dietary boron. However, although the nature and severity of the changes varied with dietary composition, many of the findings indicated that boron deprivation impairs calcium metabolism, brain function, and energy metabolism (Hunt 1994; Nielsen 1994a; Nielsen 1997). They also suggest that boron deprivation impairs immune function and exacerbates adjuvant-induced arthritis in rats.

Humans as well as animals respond to changes in dietary intake of boron (Nielsen 1994a; Penland 1994; Nielsen 1997). Among the effects of boron deprivation in men over the age of 45, postmenopausal women, and postmenopausal women on estrogen therapy were the following: an effect on macromineral and electrolyte metabolism evidenced by decreased serum 25-hydroxycholecalciferol and increased calcitonin (when dietary magnesium and copper were low); an effect on energy substrate metabolism suggested by increased serum glucose and decreased serum triglycerides; an effect on nitrogen metabolism indicated by increased blood urea nitrogen and serum creatinine and decreased urinary hydroxyproline excretion; an effect on oxidative metabolism indicated by decreased erythrocyte superoxide dismutase and serum ceruloplasmin; and an effect on erythropoiesis and hematopoiesis suggested by decreased blood hemoglobin and mean corpuscular hemoglobin content combined with increased hematocrit, platelet number, and erythrocyte number.

Boron depletion also depressed the elevation in

serum 17-beta-estradiol and plasma copper caused by estrogen ingestion, altered electroencephalograms so that they suggested diminished behavioral activation (for example, more drowsiness) and mental alertness, and impaired psychomotor skills and the cognitive processes of attention and memory.

Although a defined biochemical function is not yet known for boron, accumulated knowledge about its nutritional effects and metabolism has resulted in the recognition that boron may be of practical nutritional importance. However, no clinical disorder can be conclusively attributed to suboptimal boron nutriture. Because boron affects the macromineral metabolism, it seems likely that an inadequate boron intake may contribute to some disorders associated with suboptimal calcium metabolism, such as osteoporosis.

Chromium

Discovery of Nutritional Importance

By 1948, chromium had been recognized as a consistent component of plant and animal tissue. The first suggestion that chromium might have biologic activity appeared in 1954 (Curran 1954), when it was found that chromium enhanced the synthesis of cholesterol and fatty acids from acetate by rat liver. In 1959, trivalent chromium was identified as the active component of the “glucose tolerance factor,” which alleviated the impaired glucose tolerance in rats fed certain diets apparently inadequate in chromium (Schwarz and Mertz 1959). Between 1964 and 1968, the first reports appeared indicating that chromium could affect glucose tolerance in humans (Mertz 1993).

Subsequently, it was found that chromium supplementation also decreased serum cholesterol concentrations and normalized exaggerated insulin responses to glucose loads. Despite these suggestive findings, chromium did not receive much attention as a possible essential element for humans until 1977, when apparent chromium deficiency signs were found in a patient receiving total parenteral nutrition (Jeejeebhoy et al. 1977). Shortly thereafter, other patients receiving total parenteral nutrition were found to exhibit abnormalities of glucose metabolism that were responsive to chromium supplementation.

Dietary Importance

In 1959, it was reported that chromium-deficient rats exhibited a glucose intolerance similar to that of clinical diabetes mellitus (Schwarz and Mertz 1959). Since that time, several other deficiency signs have been described for animals, including impaired growth, elevated serum cholesterol and triglyceride concentrations, increased incidence of aortic plaques, corneal lesions, and decreased fertility and sperm count (Anderson 1988). Many of these signs were made more evident by using nutritional, metabolic, physiological, or hormonal stressors.

Signs of apparent chromium deficiency have been found in three women receiving long-term total parenteral nutrition containing low amounts of chromium. One subject, who had received total parenteral nutrition for 3.5 years, exhibited impaired glucose tolerance and glucose use, weight loss, neuropathy, elevated free fatty acid concentrations, depressed respiratory exchange ratio, and abnormalities in nitrogen metabolism (Jeejeebhoy et al. 1977). These abnormalities were alleviated by chromium supplementation. Another subject, who had received total parenteral nutrition for 5 months, developed severe glucose intolerance, weight loss, and a metabolic encephalopathy-like confusional state. All of these abnormalities were reversed by chromium supplementation (Freund, Atamian, and Fischer 1979). Chromium supplementation also reversed the development of unexplained high blood and urine glucose in a third patient who had followed a total parenteral nutrition regimen of several months' duration (Brown et al. 1986).

Although these three descriptions of human chromium deficiency are somewhat dissimilar, in all cases the apparently chromium-deficient subjects exhibited impaired glucose tolerance, or high blood glucose with glucose spilling into the urine, and a refractoriness to insulin. In addition, since 1966, a large number of reports from numerous research groups have described beneficial effects from chromium supplements in subjects with degrees of glucose intolerance ranging from low blood sugar to insulin-dependent diabetes. Most recent are reports that relatively high chromium supplementation (200 to 1000 $\mu\text{g/day}$ of chromium picolinate) decreased blood glucose and glycated hemoglobin in 162 Beijing, China, residents with high blood glucose (Anderson et al. 1997); such supplementation also decreased blood glucose, C-peptide, and insulin concentrations in gestational diabetic women undergoing a glucose tolerance test (Jovanovic-Peterson, Gutierrez, and Peterson 1996). Thus, there is a growing body of evidence suggesting that chromium supplements may be a viable treatment option for some people with diabetes resulting from inadequate synthesis of insulin or insulin resistance. In other words, chromium is potentiating the action of low amounts of insulin or improving its efficiency so that the need for exogenous sources is reduced or eliminated. Beneficial effects of chromium supplementation on blood lipid profiles also have been reported.

Even though a biochemical function has not been defined for chromium, it most likely is an essential element. Even the skeptics of chromium essentiality agree that chromium can be beneficial because of its positive effects on glucose and lipid metabolism in some individuals. This is apparently the reason that in 1980 an Estimated Safe and Adequate Daily Dietary Intake (ESADDI) was established for chromium (National Research Council 1980). For children over age 7 and adults, this amount is 50 to 200 $\mu\text{g/day}$.

There is some evidence that an intake of chromium of less than 20 $\mu\text{g/day}$ is inadequate, and dietary surveys indicate that a significant number of people consume less than this amount. Thus, it seems possible that inadequate chromium status may be partially responsible for some cases of impaired glucose tolerance, high blood glucose, low blood glucose, refractoriness to insulin, and, ultimately, diabetes. Furthermore, because impaired tissue responsiveness to insulin is a possible risk factor for cardiovascular disease, and because chromium deprivation has unfavorable effects on cholesterol and lipid metabolism, inadequate chromium status may increase susceptibility to ischemic heart disease.

Copper

Discovery of Nutritional Importance

The presence of copper in plant and animal tissues was recognized early in the nineteenth century, but this was thought to be the result of accidental contamination. The first indication that the mineral was associated with specific biochemical substances occurred when it was found in blood proteins in snails (Harless 1847). Copper was also found in the tail-feather pigment of the turaco (Church 1869) and was shown to be a constituent of octopus hemocyanin (Fredericq 1878). The discovery of reproducible amounts of copper in the human brain (Bodansky 1921) stimulated acceptance of the fact that it was definitely a physiological constituent of biological material.

Its universal distribution in plants prompted the hypothesis that copper was a catalyst participating in life processes (Fleurent and Lévi 1920; Guérithault and Maquenne 1920), and soon afterward, McHargue (1925, 1926) produced evidence suggesting that the element was beneficial in the diet of rats. Conclusive evidence that copper is an essential nutrient, however, came from studies of hemoglobin regeneration in anemic rats. In 1928, it was demonstrated that copper supplementation prevented both anemia and growth-stunting in young rats on milk-only diets (Hart et al. 1928), and subsequently, anemia caused by copper deprivation was shown to occur in chickens, cows, dogs, goats, pigs, sheep, rabbits, and humans.

Shortly after the initial discovery that rats required the mineral, certain naturally occurring copper-deficiency disorders were identified in cattle (salt-sickness in Florida and *lecksucht* in the Netherlands) and sheep (enzootic neonatal ataxia in Western Australia) (Neal, Becker, and Shealy 1931; Sjollem 1933; Bennetts and Chapman 1937). In addition to poor growth and anemia, copper deficiency was characterized by defects in pigmentation, keratinization of wool, bone formation, reproduction, myelination of the spinal cord, cardiac function, and connective tissue formation.

During the 1960s and 1970s, the first descriptions of copper deficiency in humans began to appear.

Deficiency symptoms were described for premature babies (Al-Rashid and Spangler 1971; Ashkenazi et al. 1973) and for young Peruvian and Chilean children with severe malnutrition and chronic diarrhea (Cordano, Baertl, and Graham 1964; Graham and Cordano 1969). Anemia, neutropenia, and osteoporosis were the main characteristics of the deficiency. The discovery of a genetic disease (Menkes disease) with characteristics of copper deficiency also showed that low copper status could adversely affect humans (Danks et al. 1972a, 1972b). Since 1984, reports have appeared describing signs of copper deficiency in adults brought on by depletion-repletion experiments (Klevay and Medeiros 1996).

Dietary Importance

Recent studies, mostly with rodents, have shown that copper deficiency results in a large number of defects associated with the cardiovascular system, including abnormal electrocardiograms, connective tissue abnormalities in blood vessels, enlarged hearts, altered fatty acid profiles, glucose intolerance, abnormal blood pressure, elevated blood cholesterol, pleural effusion, and ventricular aneurysms (Klevay and Medeiros 1996).

The importance of copper for malnourished children and premature infants has been described. Subsequent to its discovery, an ESADDI of 2 to 3 mg for adults was established (National Research Council 1980); some years later, this was changed to 1.5 to 3.0 mg (National Research Council 1989). Human depletion-repletion studies in adults have indicated that copper intakes of 0.65 to 1.02 mg/day are insufficient. Some subjects fed diets containing this amount exhibited changes in serum cholesterol, heart rhythm, and oxidative metabolism – measured by erythrocyte superoxide dismutase activity, serum ceruloplasmin, and mononucleated white cell cytochrome c oxidase (Klevay and Medeiros 1996). Such changes suggest that a low copper status can increase the risk for ischemic heart disease (Klevay 1990), a suggestion in urgent need of confirmation because pooled data from surveys in Belgium, Canada, the United Kingdom, and the United States have revealed that approximately one-third of 849 diets supplied copper in quantities of less than 1.0 mg/day (Klevay et al. 1993).

Fluorine

Discovery of Nutritional Importance

Fluorine first attracted nutritional attention in the 1930s, when unsuccessful attempts were made to show that low fluoride intakes produced any kind of pathology in rats and when fluoride was identified as the factor causing a mottled condition of tooth enamel known as “Colorado brown stain” and other such descriptive names (Whitford 1990). It was also noted at that time that fewer dental caries occurred in those individuals with mottled enamel, and subse-

quently it was discovered that fluoride intakes could be achieved that brought about caries reduction without mottling. As a consequence, water fluoridation was begun as a public health measure in 1945.

In the 1960s, an association was made between high fluoride intake and a reduced incidence of osteoporosis (Messer 1984). However, although the use of pharmacologic amounts of fluoride to prevent bone loss is still being investigated, its usefulness in this regard seems limited. Another round of fluoride confusion came in the early 1970s, when scientists suggested that fluoride was necessary for hematopoiesis, fertility, and growth in mice and rats (Messer 1984). It turned out, however, that this assertion was based on experiments in which animals were not fed optimal diets, and it was later concluded that the effects of fluoride were only achieved through pharmacologic mechanisms.

Dietary Importance

At present, there is no substantive evidence that fluoride is an essential nutrient. The major known biological action of fluoride is its ability to protect against pathological demineralization of calcified tissues, which is not an essential function in the true sense, but a beneficial action delivered through pharmacological mechanisms. Nonetheless, because fluoride was recognized as having beneficial properties through oral intake, ESADDIs for various times in the life cycle were established; for adults, this amount is 1.5 to 4.0 mg (National Research Council 1989).

Manganese

Discovery of Nutritional Importance

Between 1913 and 1921, manganese was found to be a constant component of plant and animal tissues, and in 1923, it was shown to be required by plants and microorganisms (McHargue 1923). Attempts to demonstrate that manganese was required by laboratory animals were unsuccessful, however, because the purified diets employed were so deficient in other essential nutrients that even the addition of manganese did not result in normal growth and survival. In 1931, it was found that manganese was essential for the growth of mice and also for normal ovarian activity in both mice and rats; it was vital to preventing testicular degeneration in rats as well (Kemmerer, Elvehjem, and Hart 1931; Orent and McCollum 1931; Waddell, Steenbock, and Hart 1931). Five years later, researchers learned that perosis (“slipped tendon”) and nutritional chondrodystrophy in poultry were caused by inadequate manganese intake (Wilgus, Norris, and Heuser 1936, 1937; Lyons and Insko 1937). The identification of a specific biochemical function for manganese remained elusive for many years, but finally, in 1962, the mineral was found to have a specific role in the synthesis of the mucopolysaccharides of cartilage (Leach and Muenster 1962). The first

report of a possible case of human manganese deficiency appeared in 1972 (Doisy 1972).

Dietary Importance

Manganese deficiency symptoms have been induced in many species of animals. They include impaired growth, skeletal abnormalities, disturbed or depressed reproductive function, ataxia of the newborn, and defects in lipid and carbohydrate metabolism (Freeland-Graves and Llanes 1994; Finley and Johnson 1996). Establishing the signs of manganese deficiency in humans, however, has been difficult.

The first possible description of manganese deficiency in humans was that of a man who had been fed a semipurified diet for an extended period of time (Doisy 1972). He developed weight loss, depressed growth of hair and nails, dermatitis, and low blood cholesterol. Moreover, his black hair developed a reddish tinge, and his clotting protein response to vitamin K supplementation was abnormal. Subsequent to the appearance of these signs, it was realized that manganese had been left out of his diet, and the subject responded to a mixed hospital diet containing manganese. Unfortunately, no supplementation with manganese alone was tried, and thus, although this case is often cited as an example of human manganese deficiency, the experimental design does not permit this to be stated conclusively.

Another report indicated that men fed only 0.11 mg/day of manganese for 35 days exhibited decreased cholesterol concentrations in serum, a fleeting dermatitis, and increased calcium, phosphorus, and alkaline phosphatase in blood (Friedman et al. 1987). Short-term manganese supplementation for 10 days, however, failed to reverse these changes, and consequently, whether the observed changes were caused by manganese deprivation can be questioned. Probably the most convincing report of a human manganese deficiency is that concerning a girl, age 4 years, who had been maintained on total parenteral nutrition that was low in manganese since she was 9 days old. She developed short stature, brittle bones, and low serum manganese; manganese supplementation improved bone density and longitudinal growth (Norose 1992). In addition to the symptoms just discussed, a low manganese status has been associated with osteoporosis and epilepsy; that is, patients with these conditions often have low serum or plasma manganese concentrations (Freeland-Graves and Turnlund 1996). Although such limited findings have made it difficult to establish the nutritional importance of manganese, most scientists do not doubt that it is essential for humans.

Molybdenum

Discovery of Nutritional Importance

From the toxicological point of view, molybdenum became important when it was discovered that the drastic scouring disease of cattle, "teart," was caused

by ingestion of excessive molybdenum, high amounts of which were present in herbage growing in the affected areas (Ferguson, Lewis, and Watson 1938). From a nutritional viewpoint, in 1930 molybdenum was reported to be essential for the growth of *Azotobacter*, a nonpathogenic bacterium found in soil and water (Bortels 1930).

Subsequently, it was discovered that molybdenum was needed by nearly all nitrogen-fixing organisms, and by 1939, the mineral was known to be both required by higher plants (independent of its role in nitrogen fixation) and consistently present in animal tissues (ter Meulen 1932; Arnon and Stout 1939). However, an early attempt to show that molybdenum was required by laboratory animals was unsuccessful (Teresi, Elvehjem, and Hart 1942), and later attempts to produce symptomatic molybdenum deficiency in rats and chicks were only successful when the diets contained massive amounts of tungsten, an antagonist of molybdenum metabolism (Higgins, Richert, and Westerfeld 1956). In 1981, the first report appeared of a human molybdenum deficiency caused by a dietary lack (Abumrad et al. 1981).

Dietary Importance

There are very few descriptions of molybdenum deficiency signs in humans (Rajagopalan 1988). In one study, four young men exhibited no clinical signs after consuming a diet with a molybdenum content of only 22 µg/day for 102 days (Turnlund et al. 1995). However, the low amount of the mineral resulted in decreased urinary excretion of uric acid and an increased urinary xanthine excretion after the administration of a load of adenosine monophosphate. These findings indicated that xanthine oxidase (a molybdoenzyme) activity was decreased.

The most convincing case of molybdenum deficiency was found in a patient receiving prolonged parenteral nutrition therapy (Abumrad et al. 1981). The patient's symptoms (which were exacerbated by methionine administration) included high blood methionine and oxypurine concentrations, low blood uric acid concentration, and low urinary excretion of uric acid and sulfate. The patient suffered mental disturbances that progressed to coma, but supplementation with ammonium molybdate improved the clinical condition, reversed the sulfur-handling defect, and normalized uric acid production.

A human genetic disorder caused by a lack of functioning molybdenum as part of the enzyme sulfite oxidase has been identified (Rajagopalan 1988). This genetic disease is characterized by severe brain damage, mental retardation, and dislocation of ocular lenses, and results in increased urinary output of sulfite, S-sulfocysteine, and thiosulfate, and a marked decrease in sulfate output.

Although molybdenum deficiency has not been recognized as a nutritional problem, it cannot be ignored because there may be unrecognized situa-

tions in which molybdenum nutrition plays an important role. For example, the molybdenum hydroxylases apparently are as important as the microsomal monooxygenase system in the metabolism of drugs and compounds foreign to the body (Beedham 1985). Perhaps low molybdenum hydroxylase activity (caused by molybdenum deficiency) would have undesirable consequences when a person is stressed by high intakes of xenobiotics. An ESADDI has been established for molybdenum; for adults it is 75 to 250 µg (National Research Council 1989).

Nickel

Discovery of Nutritional Importance

The earliest study of the biologic action of nickel appeared in 1826, when signs of oral nickel toxicity were described in rabbits and dogs (Nriagu 1980). The first reports on the presence of the mineral in plant and animal tissues appeared in 1925 (Berg 1925; Bertrand and Macheboeuf 1925), and although it was suggested that nickel was nutritionally essential in 1936 (Bertrand and Nakamura 1936), the first direct evidence for its essentiality for higher animals appeared in 1970 (Nielsen and Sauberlich 1970). By 1984, extensive signs of nickel deprivation had been reported for 6 animal species – chickens, cattle, goats, pigs, rats, and sheep (Nielsen 1984). Unfortunately, many of these symptoms may have been manifestations of pharmacologic actions of nickel; that is, high dietary nickel used in some experiments may have alleviated an abnormality caused by something other than a nutritional deficiency of nickel (for example, many diets used in the experiments were apparently low in iron) (Nielsen 1985).

Studies of nickel since 1984 probably give a more accurate picture of its nutritional role in higher animals. Support for the essentiality of the mineral has come from the identification of nickel enzymes in lower forms of life; the first to be identified was jackbean urease in 1975 (Dixon et al. 1975). Since then, nickel has been found to be an essential component of urease from bacteria, mycoplasma, fungi, yeast, algae, higher plants, and invertebrates (Mobley, Island, and Hausinger 1995). Nickel also has been identified as an essential component of three redox enzymes in bacteria involved in hydrogen oxidation, methane biogenesis, and acetate formation; these enzymes are hydrogenase, methylcoenzyme M reductase, and carbon monoxide dehydrogenase (Lancaster 1988). A nickel deficiency in humans has not been reported.

Dietary Importance

Recent studies (Nielsen 1995) suggest that nickel deprivation in goats and rats results in depressed growth, reproductive performance, and plasma glucose and alters the distribution of other elements in the body, including calcium, iron, and zinc. As with other ultratrace elements, the nature and severity of the signs

of nickel deprivation are affected by the composition of the diet. For example, both vitamin B₁₂ (Nielsen et al. 1989) and folic acid (Uthus and Poellot 1996) affect the response to nickel deprivation. This correlation, in turn, has resulted in the hypotheses that vitamin B₁₂ is necessary for the optimal expression of the biological role of nickel and that nickel has an essential function closely related to vitamin B₁₂ metabolism.

The recent discoveries of nickel essentiality and its defined biochemical functions in various microorganisms, plants, and animals indicate that it is also an essential nutrient for humans. Nonetheless, until more is known about its physiological function, it is inappropriate to specify that certain disorders might be wholly, or even partially, attributable to abnormal nickel nutrition or metabolism. But because the nickel content in some human diets can be lower than that which induces changes in animals, it may well turn out that nickel will be found to be an element of concern in human nutrition.

Selenium

Discovery of Nutritional Importance

From the biological point of view, the early history of selenium focused on its toxicological properties. The first report appeared about 1295 in the writings of Marco Polo (Trelease and Beath 1949). He described a poisonous plant (most likely seleniferous) growing in the mountains of western China, which, when eaten by beasts of burden, caused their hoofs to drop off – a phenomenon known to occur in cattle and horses that consume plants containing toxic amounts of selenium. The first scientific description of selenium toxicity, however, appeared only in 1842 (Moxon and Rhian 1943), and the element received little further attention until 1935. At that time, selenium poisoning (caused by the consumption of grains and forages high in this element) was found responsible for two diseases of livestock – known as “blind staggers” and “alkali disease” – that occurred in the northern plains of the United States (Franke and Painter 1935). This discovery stimulated investigation of the distribution of selenium in rocks, soils, plants, and animals.

In 1957, selenium came under consideration as an essential nutrient when it was found that it prevented liver necrosis in vitamin E deficient rats (Schwarz and Foltz 1957) and exudative diathesis in vitamin E deficient chicks (Patterson, Milstrey, and Stokstad 1957; Schwarz et al. 1957). The following year it was reported that muscular dystrophy, which occurred in lambs and calves in certain areas of the world, was caused by selenium deficiency (Muth et al. 1958; Proctor, Hogue, and Warner 1958). Subsequently, researchers identified areas of the world that were home to naturally occurring selenium deficiency, which impaired the growth, health, and fertility of animals living in them. Identification of a biochemical function for selenium in mammals came in 1973,

when the element was shown to be a constituent of the enzyme glutathione peroxidase (Rotruck et al. 1973). Reports first appeared in 1979 indicating that a naturally occurring clinical condition – Keshan disease – was partially attributable to low selenium intake (Keshan Disease Research Group 1979).

Dietary Importance

All known pathologies of selenium deficiency in animals can be alleviated by luxuriant intakes of vitamin E. Thus, the nutritional importance of selenium is most evident in vitamin E deficient animals. A variety of pathological changes attributable to combined deficiencies of selenium and vitamin E have been described for several animal species. These changes result in dysfunctions of the brain, the cardiovascular system, the liver, the muscles, and the fetus. Many of the pathological conditions caused by the combined deficiencies can be modified by other dietary factors, including the intake of polyunsaturated fatty acids, sulfur amino acids, and synthetic antioxidants (Combs and Combs 1984).

Evidence that selenium can be of practical nutritional importance came from studies of the relationship between selenium and Keshan disease, an endemic cardiomyopathy affecting children and young women that occurs in an area running from northeastern to southwestern China (Levander and Burk 1994). The primary pathological changes in Keshan disease include heart enlargement with multiple focal degenerative necrosis and fibrous replacement of the myocardium. Selenium supplementation was found effective in preventing Keshan disease, the incidence of which varied with several other factors, and it has been suggested that low vitamin E status or oxidant stress caused by viral or bacterial infection may be contributory factors. Other reports of selenium deficiency came from studies of patients on long-term total parenteral nutrition. The signs of deficiency in these patients included bilateral muscular discomfort and a dilated cardiomyopathy that histopathologically resembled Keshan disease (Levander and Burk 1994).

Finally, it was found that a luxuriant intake of selenium achieved by supplementation (200 µg/day) was effective in preventing the occurrence of some forms of cancer (Clark et al. 1996). Because the preceding discovery indicated that those individuals who are selenium deficient can be considered at risk for developing pathology, people should achieve an intake meeting the Recommended Dietary Allowance of 70 and 55 µg/day for adult men and women, respectively (National Research Council 1989).

Silicon

Discovery of Nutritional Importance

Silicon has long been thought important in maintaining health in humans. Before much was known about it in biology, one of the luminaries of medical science,

Louis Pasteur, predicted that silicon would prove to be an important therapeutic substance for many diseases (Becker et al. 1983), and at the beginning of the twentieth century, numerous French and German reports suggested that the prediction of Pasteur would become fact. These reports described therapeutic successes in treating numerous diseases – including atherosclerosis, hypertension, and dermatitis – with sodium silicate, with simple organic silicon compounds, and with tea made from the silicon-rich horsetail plant. However, by 1930, these “therapeutic successes” had come to be viewed as failures, and in the face of inadequate evidence for silicon’s biological activity, the mineral faded into medical obscurity.

During the next four decades, silicon – as consumed in the diet – was generally considered a biologically inert, harmless, nonessential element for living organisms, except for some lower forms of life (diatoms, radiolarians, and sponges) in which silica meets a structural need (Carlisle 1984). But then, in 1972, it was reported that silicon was essential for bone formation (Carlisle 1972), and at about the same time, other studies appeared suggesting, as had those of the early twentieth century, that inadequate dietary silicon may contribute to some cases of atherosclerosis, hypertension, some bone disorders, and the aging process. However, these studies have generally been ignored because a specific biochemical role for silicon, which could explain the mechanism behind such occurrences, has not been identified.

Dietary Importance

Most of the signs of silicon deficiency in animals indicate aberrant metabolism of connective tissue and bone. Illustrative are chicks on a silicon-deficient diet that exhibit structural abnormalities of the skull, depressed collagen content in bone, and long-bone abnormalities characterized by small, poorly formed joints and defective endochondral bone growth (Carlisle 1984). Silicon deprivation can also affect the response to other dietary manipulations. Thus, although rats fed a diet low in calcium and high in aluminum accumulated high amounts of aluminum in the brain, it was found that relatively high silicon supplements prevented such accumulation (Carlisle and Curran 1987).

Similarly, whereas high dietary aluminum depressed brain zinc concentrations in thyroidectomized rats fed low dietary silicon, relatively high silicon supplements prevented the depression (Carlisle, Curran, and Duong 1991). This effect, however, was not seen in nonthyroidectomized rats. Other biochemical actions indicating that silicon is an essential element include its consistent presence in collagen; and in bone tissue culture it is needed for maximal bone prolylhydroxylase activity (Carlisle, Berger, and Alpenfels 1981). Silicon deficiency decreases ornithine aminotransferase (an enzyme in the collagen formation pathway) in rats (Seaborn and Nielsen 1996). Thus, ample evidence

exists to suggest that silicon is an essential nutrient for higher animals, including humans.

Findings from animals indicating that silicon nutriture affects macromolecules – such as glycosaminoglycans, collagen, and elastin – suggest that silicon is needed for healthy bones, brains, and blood vessels (Seaborn and Nielsen 1993). Although more needs to be known about the physiological or biochemical function and requirement for silicon, speculation has materialized on the possible involvement of silicon deprivation in the occurrence of several human disorders, including atherosclerosis, osteoarthritis, osteoporosis, hypertension, and Alzheimer's disease. Such speculation indicates the need for more work to clarify the consequences of silicon deficiency in humans.

Vanadium

Discovery of Nutritional Importance

A report on the toxicity of vanadium was published in 1876 (Priestley and Gamgee 1876) but apparently failed to stimulate interest in the mineral's biological actions. Consequently, the subject lay dormant until the turn of the twentieth century, when various French physicians used vanadium as a panacea for a number of human disorders. Shortly thereafter, a classic paper on the pharmacological and toxicological actions of vanadium appeared (Jackson 1912) – at about the time that extremely high vanadium concentrations were found in the blood of ascidian worms (Henze 1911, 1912). The hypothesis that vanadium has a physiological role in higher animals has had a long, yet muddled, history. In 1949, O. Rygh reported that vanadium might be needed by animals because it markedly stimulated the mineralization of bones and teeth and prevented caries formation in rats and guinea pigs. In 1950, D. Bertrand wrote that “we are completely ignorant of the physiological role of vanadium in animals, where its presence is constant.” In 1963, H. A. Schroeder, J. J. Balassa, and I. H. Tipton indicated that although vanadium behaves like an essential trace metal, final proof of its essentiality for mammals was still lacking. But in 1974, L. L. Hopkins, Jr., and H. E. Mohr stated: “[W]e are secure in the concept that vanadium is an essential nutrient.” Subsequent reviews, however, have presented a convincing argument that the evidence for the nutritional essentiality of vanadium was inconclusive (Nielsen 1984, 1985). That is, much of the evidence for essentiality may have been the result of high vanadium supplements (10 to 100 times the amount normally found in natural diets) inducing pharmacologic changes in animals fed imbalanced diets. The most substantive evidence for vanadium essentiality has appeared only since 1987.

Dietary Importance

Signs of apparent vanadium deprivation have been described in goats (Anke et al. 1989) and rats (Uthus and Nielsen 1990; Nielsen, Poellot, and Uthus 1997).

For goats, such signs include depressed milk production and skeletal deformations of the forelegs, including thickening of the forefoot tarsal joints. Vanadium-deprived rats exhibit altered thyroid hormone metabolism, especially if stressed with high or low iodine intakes. Supporting the suggestion that vanadium is an essential element for higher animals are the findings of functional roles for vanadium in lower forms of life. Algae, lichens, fungi, and bacteria all have enzymes that require vanadium for activity. These enzymes include haloperoxidases, which catalyze the oxidation of halide ions by hydrogen peroxide, thus facilitating the formation of a carbon-hydrogen bond (Vilter 1995).

Numerous biochemical and physiological functions for vanadium have been suggested on the basis of its in-vitro actions on cells and pharmacologic actions in animals; these actions (Willsky 1990) include insulin-mimetic properties, numerous stimulatory effects on cell proliferation and differentiation, effects of cell phosphorylation-dephosphorylation, effects on glucose and ion transport across the plasma membrane, and effects on oxidation-reduction processes.

Based on circumstantial evidence from animal deprivation studies and the functional roles of vanadium in lower forms of life, it seems quite possible that the mineral may be found essential for some enzyme reaction important to human health, perhaps one that is involved in thyroid metabolism. However, until a biochemical function is definitively identified for vanadium, its nutritional importance will be uncertain. Knowledge of a biochemical function is also needed in order to disentangle the element's pharmacologic actions from its nutritional actions. Yet, because vanadium is so pharmacologically active, a beneficial pharmaceutical role for this element may be found.

Summary

Trace elements were known to be present in biological materials more than 100 years ago, and several were found to be essential for plant and animal life more than 60 years ago. Studies of the pharmacological and beneficial properties of various trace elements, which were being done in the nineteenth century, continue today. However, except for iron, iodine, and cobalt (as cobalamin), the possibility that some trace elements are of practical nutritional importance did not receive much attention until about 30 years ago. At that time, excitement abounded about trace elements, and the new knowledge of them was considered a breakthrough toward better health.

One trace element after another was identified as essential, or at least suggested to be essential, and new functions were being found for those already established as essential. As these new essential elements or new functions were identified, hopes were raised that trace element nutriture would help to unravel the

causes of some diseases, especially those that are chronic and associated with aging. Unfortunately, this has not occurred, but research does continue to show that some trace elements are more important than is currently acknowledged. The history of investigations into the trace elements iron, iodine, zinc, and selenium indicates a likelihood that several more elements – especially boron, copper, and chromium – will be found to be of nutritional (and clinical) importance for human health and longevity.

Forrest H. Nielsen

Bibliography

- Abumrad, N. N., A. J. Schneider, D. Steel, and L. S. Rogers. 1981. Amino acid intolerance during prolonged total parenteral nutrition reversed by molybdate therapy. *American Journal of Clinical Nutrition* 34: 2551-9.
- Al-Rashid, R. A., and J. Spangler. 1971. Neonatal copper deficiency. *New England Journal of Medicine* 285: 841-3.
- Anderson, R. A. 1988. Chromium. In *Trace minerals in foods*, ed. K. T. Smith, 231-47. New York.
- Anderson, R. A., N. Cheng, N. Bryden, et al. 1997. Elevated intakes of supplemental chromium improve glucose and insulin variables with type II diabetes. *Diabetes* 46: 1786-91.
- Anke, M. 1986. Arsenic. In *Trace elements in human and animal nutrition*, Vol. 2, ed. W. Mertz, 347-72. Orlando, Fla.
- Anke, M., B. Groppel, K. Gruhn, et al. 1989. The essentiality of vanadium for animals. In *6th International Trace Element Symposium*, ed. M. Anke, W. Baumann, H. Bräunlich, et al., 17-27. Jena, Germany.
- Arnon, D. I., and P. R. Stout. 1939. Molybdenum as an essential element for higher plants. *Plant Physiology* 14: 599-602.
- Ashkenazi, A., S. Levin, M. Djaldetti, et al. 1973. The syndrome of neonatal copper deficiency. *Pediatrics* 52: 525-33.
- Becker, C.-H., D. Matthias, H. Wossmann, et al. 1983. Investigations on a possible medical importance of silicon. In *4. Spurenelement Symposium Jena*, ed. M. Anke, W. Baumann, H. Bräunlich, and C. Brückner, 142-8. Jena, Germany.
- Beedham C. 1985. Molybdenum hydroxylases as drug-metabolizing enzymes. *Drug Metabolism Reviews* 16: 119-56.
- Bennetts, H. W., and F. E. Chapman. 1937. Copper deficiency in sheep in Western Australia: A preliminary account of the etiology of enzootic ataxia of lambs and an anemia of ewes. *Australian Veterinary Journal* 13: 138-49.
- Berg, R. 1925. Das Vorkommen seltener Elemente in den Nahrungsmitteln und menschlichen Ausscheidungen. *Biochemische Zeitschrift* 165: 461-2.
- Bertrand, D. 1950. Survey of contemporary knowledge of biogeochemistry. 2. The biogeochemistry of vanadium. *Bulletin of the American Museum of Natural History* 94: 403-56.
- Bertrand, G., and M. Mâcheboeuf. 1925. Sur la présence du nickel et du cobalt chez les animaux. *Comptes Rendus de l'Académie des Sciences* 180: 1380-3.
- Bertrand, G., and H. Nakamura. 1936. Recherches sur l'importance physiologique du nickel et du cobalt. *Bulletin de la Société Scientifique d'Hygiène Alimentaire et d'Alimentation Rationnelle* 24: 338-43.
- Bodansky, M. 1921. The zinc and copper content of the human brain. *Journal of Biological Chemistry* 48: 361-4.
- Bortels, H. 1930. Molybdenum as a catalyst in the biological fixation of nitrogen. *Archiv für Mikrobiologie* 1: 333-42.
- Brown, R. O., S. Forloines-Lynn, R. E. Cross, and W. D. Heizer. 1986. Chromium deficiency after long-term total parenteral nutrition. *Digestive Diseases and Sciences* 31: 661-4.
- Carlisle, E. M. 1972. Silicon: An essential element for the chick. *Science* 178: 619-21.
1984. Silicon. In *Biochemistry of the essential ultratrace elements*, ed. E. Frieden, 257-91, New York.
- Carlisle, E. M., J. W. Berger, and W. F. Alpenfels. 1981. A silicon requirement for prolylhydroxylase activity. *Federation Proceedings* 40: 886.
- Carlisle, E. M., and M. J. Curran. 1987. Effect of dietary silicon and aluminum on silicon and aluminum levels in rat brain. *Alzheimer Disease and Associated Disorders* 1: 83-9.
- Carlisle, E. M., M. J. Curran, and T. Duong. 1991. The effect of interrelationships between silicon, aluminum and the thyroid on zinc content in the brain. In *Trace elements in man and animals - 7*, ed. B. Momčilović, 12.16-17. Zagreb, Croatia.
- Church, A. H. 1869. Researches on turacin, an animal pigment containing copper. *Philosophical Transactions of the Royal Society of London* 159: 627.
- Clark, L. C., G. F. Combs, Jr., B. W. Turnbull, et al. 1996. Effects of selenium supplementation for cancer prevention in patients with carcinoma of the skin: A randomized controlled trial. *Journal of the American Medical Association* 276: 1957-63.
- Combs, G. F., Jr., and S. B. Combs. 1984. The nutritional biochemistry of selenium. *Annual Review of Nutrition* 4: 257-80.
- Cordano, A., J. M. Baertl, and G. G. Graham. 1964. Copper deficiency in infancy. *Pediatrics* 34: 324-6.
- Curran, G. L. 1954. Effect of certain transition group elements on hepatic synthesis of cholesterol in the rat. *Journal of Biological Chemistry* 210: 765-70.
- Danks, D. M., P. E. Campbell, B. J. Stevens, et al. 1972a. Menkes' kinky hair syndrome. An inherited defect in copper absorption with widespread effects. *Pediatrics* 50: 188-201.
- Danks, D. M., B. J. Stevens, P. E. Campbell, et al. 1972b. Menkes' kinky-hair syndrome. *Lancet* 1: 1100-3.
- De Renzo, E. C., E. Kaleita, P. G. Heytler, et al. 1953. Identification of the xanthine oxidase factor as molybdenum. *Archives of Biochemistry and Biophysics* 45: 247-53.
- Dixon, N. E., C. Gazzola, R. L. Blakeley, and B. Zerner. 1975. Jack bean urease (EC 3.5.1.5). A metalloenzyme. A simple biological role for nickel? *Journal of the American Chemical Society* 97: 4131-3.
- Doisy, E. A., Jr. 1972. Micronutrient controls on biosynthesis of clotting proteins and cholesterol. In *Trace substances in environmental health*, ed. D. D. Hemphill, 193-9. Columbia, Mo.
- Ferguson, W. S., A. H. Lewis, and S. J. Watson. 1938. Action of molybdenum in nutrition of milking cattle. *Nature* 141: 553.
- Finley, J. W., and P. E. Johnson. 1996. Manganese deficiency and excess in rodents. In *Trace elements in laboratory rodents*, ed. R. R. Watson, 85-106. Boca Raton, Fla.
- Fleurent, E., and L. Lévi. 1920. Sur la présence du cuivre dans l'organisme végétal et animal. *Bulletin de la Société Chimique France* 27: 441-2.

- Follis, R. H., Jr. 1947. The effect of adding boron to a potassium-deficient diet in the rat. *American Journal of Physiology* 150: 520-2.
- Franke, K. W., and E. P. Painter. 1935. Selenium in proteins from toxic foodstuffs. IV. The effect of feeding toxic proteins, toxic protein hydrolysates, and toxic protein hydrolysates from which the selenium has been removed. *Journal of Nutrition* 10: 599-611.
- Fredericq, L. 1878. Recherches sur la physiologie du poulpe commun (*Octopus vulgaris*). *Archives de Zoologie Expérimentale et Générale* 7: 535.
- Freeland-Graves, J., and C. Llanes. 1994. Models to study manganese deficiency. In *Manganese in health and disease*, ed. D. J. Klimis-Tavantzis, 59-86. Boca Raton, Fla.
- Freeland-Graves, J. H., and J. R. Turnlund. 1996. Deliberations and evaluations of the approaches, endpoints and paradigms for manganese and molybdenum dietary recommendations. *Journal of Nutrition* 126: 2435S-40S.
- Freund, H., S. Atamian, and J. E. Fischer. 1979. Chromium deficiency during total parenteral nutrition. *Journal of the American Medical Association* 241: 496-8.
- Friedman, B. J., J. Freeland-Graves, C. Bales, et al. 1987. Manganese balance and clinical observations in young men fed a manganese-deficient diet. *Journal of Nutrition* 117: 133-43.
- Gordon, V., ed. 1987. The case of the toxic life-preserver. *Borax Review* 2: 10-12.
- Graham, G. G., and A. Cordano. 1969. Copper depletion and deficiency in the malnourished infant. *Jobns Hopkins Medical Journal* 124: 139-50.
- Guérithault, B., and M. L. Maquenne. 1920. Sur la présence du cuivre dans les plantes et particulièrement dans les matières alimentaires d'origine végétal. *Comptes Rendus des Sciences Société de Biologie et de ses Filiales et Associées* 171: 196-8.
- Harless, E. 1847. Über das blaue Blut einiger wirbellosen Thiere und dessen Kupfergehalt. *Müller's Archiv für Anatomie und Physiologie*, 148-56.
- Hart, E. B., H. Steenbock, J. Waddell, and C. A. Elvehjem. 1928. Iron in nutrition. VII. Copper as a supplement to iron for hemoglobin building in the rat. *Journal of Biological Chemistry* 77: 797-812.
- Henze, M. 1911. Untersuchungen über das Blut der Ascidien. I. Mitteilung. Die Vanadiumverbindung der Blutkörperchen. *Hoppe-Seyler's Zeitschrift Physiologische Chemistry* 72: 494-501.
1912. Untersuchungen über das Blut der Ascidien. *Hoppe-Seyler's Zeitschrift Physiologische Chemistry* 79: 215-28.
- Higgins, E. S., D. A. Richert, and W. W. Westerfeld. 1956. Molybdenum deficiency and tungstate inhibition studies. *Journal of Nutrition* 59: 539-59.
- Hopkins, L. L., Jr., and H. E. Mohr. 1974. Vanadium as an essential nutrient. *Federation Proceedings* 33: 1773-5.
- Hove, E., C. A. Elvehjem, and E. B. Hart. 1938. Arsenic in the nutrition of the rat. *American Journal of Physiology* 124: 205-12.
1939. Boron in animal nutrition. *American Journal of Physiology* 127: 689-701.
- Hunt, C. D. 1994. The biochemical effects of physiologic amounts of dietary boron in animal nutrition models. *Environmental Health Perspectives* 102 (Supplement 7): 35-43.
- Hunt, C. D., and F. H. Nielsen. 1981. Interaction between boron and cholecalciferol in the chick. In *Trace elements in man and animals (TEMA-4)*, ed. J. McC. Howell, J. M. Gawthorne, and C. L. White, 597-600. Canberra, Australia.
- Jackson, D. E. 1912. The pharmacological action of vanadium. *Journal de Pharmacologie* 3: 477-514.
- Jeejeebhoy, K. N., R. C. Chu, E. B. Marliss, et al. 1977. Chromium deficiency, glucose intolerance, and neuropathy reversed by chromium supplementation, in a patient receiving long-term total parenteral nutrition. *American Journal of Clinical Nutrition* 30: 531-8.
- Jovanovic-Peterson, L., M. Gutierrez, and C. M. Peterson. 1996. Chromium supplementation for gestational diabetic women (GDM) improves glucose tolerance and decreases hyperinsulinemia. *Diabetes* 45: Abs 1263.
- Kemmerer, A. R., C. A. Elvehjem, and E. B. Hart. 1931. Studies on the relation of manganese to the nutrition of the mouse. *Journal of Biological Chemistry* 92: 623-30.
- Keshan Disease Research Group. 1979. Observation on the effect of sodium selenite in prevention of Keshan disease. *Chinese Medical Journal* 92: 471-82.
- Klevay, L. M. 1990. Ischemic heart disease: Toward a unified theory. In *Role of copper in lipid metabolism*, ed. K. Y. Lei and T. P. Carr, 233-67. Boca Raton, Fla.
- Klevay, L. M., J. P. Buchet, V. W. Bunker, et al. 1993. Copper in the western diet (Belgium, Canada, U.K., and USA). In *Trace elements in man and animals - TEMA-8*, ed. M. Anke, D. Meissner, and C. F. Mills, 207-10. Gersdorf, Germany.
- Klevay, L. M., and D. M. Medeiros. 1996. Deliberations and evaluations of the approaches, endpoints and paradigms for dietary recommendations about copper. *Journal of Nutrition* 126: 2419S-26S.
- Lancaster, J. R., Jr., ed. 1988. *The bioinorganic chemistry of nickel*. New York.
- Leach, R. M., Jr., and A.-M. Muenster. 1962. Studies on the role of manganese in bone formation. I. Effect upon the mucopolysaccharide content of chick bone. *Journal of Nutrition* 78: 51-6.
- Levander, O. A., and R. F. Burk. 1994. Selenium. In *Modern nutrition in health and disease*. Eighth edition, Vol. 1, ed. M. E. Shils, J. A. Olson, and M. Shike, 242-51. Philadelphia, Pa.
- Lyons, M., and W. M. Insko, Jr. 1937. Chondrodystrophy in the chick embryo produced by manganese deficiency in the diet of the hen. *Kentucky Agricultural Experimental Station Bulletin* 371: 61-75.
- Mayer, D. R., W. Kosmus, H. Poggitsch, et al. 1993. Essential trace elements in humans. Serum arsenic concentrations in hemodialysis patients in comparison to healthy controls. *Biological Trace Element Research* 37: 27-38.
- McHargue, J. S. 1923. Effect of different concentrations of manganese sulfate on the growth of plants in acid and neutral soils and the necessity of manganese as a plant nutrient. *Journal of Agricultural Research* 24: 781-93.
1925. The association of copper with substances containing the fat-soluble A vitamin. *American Journal of Physiology* 72: 583-94.
1926. Further evidence that small quantities of copper, manganese and zinc are factors in the metabolism of animals. *American Journal of Physiology* 77: 245-55.
- Mertz, W. 1993. Chromium in human nutrition: A review. *Journal of Nutrition* 123: 626-33.
- Messer, H. H. 1984. Fluorine. In *Biochemistry of the essential ultratrace elements*, ed. E. Frieden, 55-87. New York.
- Mobley, H. L., M. D. Island, and R. P. Hausinger. 1995. Molecular biology of microbiological ureases. *Microbiological Reviews* 59: 451-80.
- Moxon, A. L., and M. Rhian. 1943. Se poisoning. *Physiological Reviews* 23: 305-37.
- Muth, O. H., J. E. Oldfield, L. F. Remmert, and J. R. Schubert.

1958. Effects of selenium and vitamin E on white muscle disease. *Science* 128: 1090.
- National Research Council. 1980. *Recommended dietary allowances*. Ninth edition. Washington, D.C.
1989. *Recommended dietary allowances*. Tenth edition. Washington, D.C.
- Neal, W. M., R. B. Becker, and A. L. Shealy. 1931. A natural copper deficiency in cattle rations. *Science* 74: 418-19.
- Nielsen, F. H. 1980. Interactions between essential trace and ultratrace elements. *Annals of the New York Academy of Sciences* 355: 152-64.
1984. Ultratrace elements in nutrition. *Annual Review of Nutrition* 4: 21-41.
1985. The importance of diet composition in ultratrace element research. *Journal of Nutrition* 115: 1239-47.
- 1994a. Biochemical and physiologic consequences of boron deprivation in humans. *Environmental Health Perspectives* 102 (Supplement 7): 59-63.
- 1994b. Ultratrace minerals. In *Modern nutrition in health and disease*. Eighth edition, ed. M. E. Shils, J. A. Olson, and M. Shike, 269-86. Philadelphia, Pa.
1995. Individual functional roles of metal ions *in vivo*. Beneficial metal ions. Nickel. In *Handbook of metal-ligand interactions in biological fluids*. *Bioinorganic medicine*, Vol. 1, 257-60. New York.
1997. Boron in human and animal nutrition. *Plant and Soil* 193: 199-208.
- Nielsen, F. H., R. A. Poellot, and E. O. Uthus. 1997. Vanadium deprivation alters the changes in plasma thyroid hormone concentrations, pancreatic amylase activity, and serum lactate dehydrogenase activity in the BB rat caused by increasing dietary iodine. *FASEB Journal* 11: A148.
- Nielsen, F. H., and H. E. Sauberlich. 1970. Evidence of a possible requirement for nickel by the chick. *Proceedings of the Society for Experimental Biology and Medicine* 134: 845-9.
- Nielsen, F. H., and E. O. Uthus. 1984. Arsenic. In *Biochemistry of the essential ultratrace elements*, ed. E. Frieden, 319-40. New York.
- Nielsen, F. H., T. J. Zimmerman, T. R. Shuler, et al. 1989. Evidence for a cooperative metabolic relationship between nickel and vitamin B₁₂ in rats. *Journal of Trace Elements in Experimental Medicine* 2: 21-9.
- Norose, N. 1992. Manganese deficiency in a child with very short bowel syndrome receiving long-term parenteral nutrition. *Journal of Trace Elements in Experimental Medicine* 5: 100-1.
- Nriagu, Jerome O. ed. 1980. *Nickel in the environment*. New York.
- Orent, E. R., and E. V. McCollum. 1931. Effects of deprivation of manganese in the rat. *Journal of Biological Chemistry* 92: 651-78.
- Orent-Keiles, E. 1941. The role of boron in the diet of the rat. *Proceedings of the Society for Experimental Biology and Medicine* 44: 199-202.
- Patterson, E. L., R. Milstrey, and E. L. R. Stokstad. 1957. Effect of selenium in preventing exudative diathesis in chicks. *Proceedings of the Society for Experimental Biology and Medicine* 95: 617-20.
- Penland, J. G. 1994. Dietary boron, brain function, and cognitive performance. *Environmental Health Perspectives* 102 (Supplement 7): 65-72.
- Priestley, J., and A. Gamgee. 1876. On the physiological action of vanadium. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* 166: 495-556.
- Proctor, J. F., D. E. Hogue, and R. G. Warner. 1958. Selenium, vitamin E and linseed oil meal as preventatives of muscular dystrophy in lambs. *Journal of Animal Science* 17: 1183-4.
- Rajagopalan, K. V. 1988. Molybdenum: An essential trace element in human nutrition. *Annual Review of Nutrition* 8: 401-27.
- Richert, D. A., and W. W. Westerfeld. 1953. Isolation and identification of the xanthine oxidase factor as molybdenum. *Journal of Biological Chemistry* 203: 915-23.
- Rotruck, J. T., A. L. Pope, H. E. Ganther, et al. 1973. Selenium: Biochemical role as a component of glutathione peroxidase. *Science* 179: 588-90.
- Rygh, O. 1949. Recherches sur les oligo-éléments. II. De l'importance du thallium et du vanadium, du silicium et du fluor. *Bulletin de la Société Chimie Biologique* 31: 1403-7.
- Schroeder, H. A., and J. J. Balassa. 1966. Abnormal trace metals in men: Arsenic. *Journal of Chronic Diseases* 19: 85-106.
- Schroeder, H. A., J. J. Balassa, and I. H. Tipton. 1963. Abnormal trace metals in man - vanadium. *Journal of Chronic Diseases* 16: 1047-71.
- Schwarz, K., J. G. Bieri, G. M. Briggs, and M. L. Scott. 1957. Prevention of exudative diathesis in chicks by factor 3 and selenium. *Proceedings of the Society for Experimental Biology and Medicine* 95: 621-5.
- Schwarz, K., and C. M. Foltz. 1957. Selenium as an integral part of factor 3 against dietary necrotic liver degeneration. *Journal of the American Chemical Society* 79: 3292-3.
- Schwarz, K., and W. Mertz. 1959. Chromium (III) and the glucose tolerance factor. *Archives of Biochemistry and Biophysics* 85: 292-5.
- Seaborn, C. D., and F. H. Nielsen. 1993. Silicon: A nutritional beneficence for bones, brains and blood vessels? *Nutrition Today* 28: 13-18.
1996. Dietary silicon affects some collagen synthesizing enzymes in rats. *FASEB Journal* 10: A784.
- Sjollema, B. 1933. Kupfermangel als Ursache von Krankheiten bei Pflanzen und Tieren. *Biochemische Zeitschrift* 267: 151-6.
- Skinner, J. T., and J. S. McHargue. 1945. Response of rats to boron supplements when fed rations low in potassium. *American Journal of Physiology* 143: 385-90.
1946. Supplementary effects of arsenic and manganese on copper in the synthesis of hemoglobin. *American Journal of Physiology* 145: 500-6.
- Sommer, A. L., and C. B. Lipman. 1926. Evidence of the indispensable nature of zinc and boron for higher green plants. *Plant Physiology* 1: 231-49.
- Teresi, J. D., C. A. Elvehjem, and E. B. Hart. 1942. Molybdenum in the nutrition of the rat. *American Journal of Physiology* 137: 504-8.
- Teresi, J. D., E. Hove, C. A. Elvehjem, and E. B. Hart. 1944. Further study of boron in the nutrition of rats. *American Journal of Physiology* 140: 513-18.
- ter Meulen, H. 1932. Distribution of molybdenum. *Nature* 130: 966.
- Trelease, S. F., and O. A. Beath. [1892] 1949. *Selenium. Its geological occurrence and its biological effects in relation to botany, chemistry, agriculture, nutrition, and medicine by Sam F. Trelease and Orville A. Beath*. New York and Laramie, Wyo.
- Turnlund, J. R., W. R. Keyes, G. L. Peiffer, and G. Chiang. 1995. Molybdenum absorption, excretion and retention studied with stable isotopes in young men during depletion and repletion. *American Journal of Clinical Nutrition* 61: 1102-9.
- Uthus, E. O. 1994. Arsenic essentiality and factors affecting its importance. In *Arsenic. Exposure and health*, ed. W. R. Chappell, C. O. Abernathy, and C. R. Cothorn, 199-208. Northwood, England.

- Uthus, E. O., and F. H. Nielsen. 1990. Effect of vanadium, iodine and their interaction on growth, blood variables, liver trace elements and thyroid status indices in rats. *Magnesium and Trace Elements* 9: 219–26.
- Uthus, E. O., and R. A. Poellot. 1996. Dietary folate affects the response of rats to nickel deprivation. *Biological Trace Element Research* 52: 23–35.
- Vilter, H. 1995. Vanadium-dependent haloperoxidases. In *Metal ions in biological systems, Vol. 31: Vanadium and its role in life*, ed. H. Sigel and A. Sigel, 325–62. New York.
- Waddell, J., H. Steenbock, and E. B. Hart. 1931. Growth and reproduction on milk diets. *Journal of Nutrition* 4: 53–65.
- Warrington, K. 1923. The effect of boric acid and borax on the broad bean and certain other plants. *Annals of Botany* 37: 629–72.
- Whitford, G. M. 1990. The physiological and toxicological characteristics of fluoride. *Journal of Dental Research* 69: 539–49.
- Wiley, H. W. 1904. Influence of food preservatives and artificial colors on digestion and health. I. Boric acid and borax. *U.S. Department of Agriculture Bulletin No. 84*. Washington, D.C.
- Wilgus, H. S., Jr., L. C. Norris, and G. F. Heuser. 1936. The role of certain inorganic elements in the cause and prevention of perosis. *Science* 84: 252–3.
1937. The role of manganese and certain other trace elements in the prevention of perosis. *Journal of Nutrition* 14: 155–67.
- Willsky, G. R. 1990. Vanadium in the biosphere. In *Vanadium in biological systems. Physiology and biochemistry*, ed. N. D. Chasteen, 1–24. Dordrecht, the Netherlands.

IV.B.9 🐉 Zinc

In 1869, J. Raulin showed for the first time that zinc is a growth factor for *Aspergillus niger*. Then, in 1926, it was discovered that zinc is essential for higher plants (Sommer and Lipman 1926). The importance of zinc in the growth and development of rats was demonstrated in 1934 (Todd, Elvehjem, and Hart 1934), and in 1955, H. F. Tucker and W. D. Salmon related a disease in swine called parakeratosis to a deficiency of zinc. Shortly thereafter, zinc was shown to be a growth factor for chickens (O'Dell, Newberne, and Savage 1958).

The manifestations of zinc deficiency in animals include growth failure, loss of hair, thickening and hyperkeratinization of the epidermis, and testicular atrophy. Zinc deficiency in breeding hens results in decreased hatchability and gross anomalies in embryonic skeletal development.

Although the importance of zinc for animals was established 60 years ago, it has only been during the past 30 years that zinc deficiency in humans has been recognized. In 1974, the Food and Nutrition Board of the National Research Council of the National Academy of Sciences made a landmark decision in establishing a recommended dietary allowance (RDA) for zinc.

Discovery of Zinc Deficiency in Humans

Studies in Iran

In 1958, this author joined the staff of Dr. Hobart A. Reimann, Chief of Medicine at the Nemazee Hospital of Pahlavi University in Shiraz, Iran. In the fall of that year, Dr. James A. Halsted of the Saadi Hospital of Pahlavi University invited me to discuss a patient who had severe anemia.

The patient was a 21-year-old male, who appeared to be only a 10-year-old boy. In addition to severe growth retardation and anemia, he had hypogonadism, hepatosplenomegaly, rough and dry skin, mental lethargy, and geophagia. His intake of animal protein had been negligible. He ate only wheat flour and unleavened bread and consumed nearly 0.5 kilograms (kg) of clay daily (the habit of geophagia in the villages around Shiraz is fairly common). Ten more cases that were similar arrived at the hospital for my care within a short period of time.

We documented the existence of iron-deficiency anemia in the patients, but there was no evidence of blood loss. We considered three possible mechanisms of iron deficiency: (1) The availability of iron in the high-cereal, protein-containing diet was most probably very low due to high phytate levels in the bread, which bind iron; (2) that geophagia further decreased iron absorption (as Minnich and colleagues [1968] observed later); and (3) that an excessive loss of iron by sweating in the hot summers of Iran may have contributed significantly to negative iron balance.

After administration of ferrous sulfate (1 gram per day) orally and a nutritious hospital diet containing adequate animal protein, the anemia was corrected, hepatosplenomegaly improved, subjects grew pubic hair, and genitalia size increased (Prasad, Halsted, and Nadimi 1961). Liver function tests were unremarkable except for the serum alkaline phosphatase activity, which increased after treatment. Retrospectively, two explanations seem plausible: (1) Ordinary pharmaceutical preparations of iron might have contained appreciable quantities of zinc as a contaminant, and (2) animal protein in the diet most likely supplied available zinc, thus inducing the activity of alkaline phosphatase, a known zinc metalloenzyme.

It was difficult to account for all of the clinical features solely by tissue iron deficiency because growth retardation and testicular atrophy are not normally found in iron-deficient experimental animals. Moreover, since this syndrome was fairly prevalent in the villages near Shiraz, the rare syndrome of hypopituitarism as an explanation for growth retardation and hypogonadism was considered to be very unlikely.

We explored the possibility that zinc deficiency may have been present concomitantly with iron deficiency in these patients (Prasad et al. 1961). Because heavy metals may form insoluble complexes with phosphate, we speculated that some factors responsible for decreased availability of iron in these patients

with geophagia may also have decreased the availability of zinc. B. L. O'Dell and J. E. Savage (1960) observed that the phytate (inositol hexaphosphate) present in cereal grains markedly impaired the absorption of zinc. Changes in the activity of alkaline phosphatase as observed in our patients had also been noticed after the zinc supplementation of deficient animals. Thus, it seemed that the dwarfism, testicular atrophy, retardation of skeletal maturation, and changes in serum alkaline phosphatase activity of our subjects might be explained by zinc deficiency.

I. I. Lemann (1910) had previously reported a similar clinical syndrome in patients who had hookworm infection in the United States, but this was not related to a nutritional deficiency. Similar cases from Turkey were reported by F. Reimann (1955), but without detailed descriptions. He considered a genetic defect to be a possible explanation for certain aspects of the clinical syndrome.

Studies in Egypt

In October 1960, Dr. William J. Darby invited me to meet with him at the U.S. Naval Medical Research Unit in Cairo, Egypt, where I shared with him my speculation that zinc deficiency in the Middle East was prevalent and was responsible for widespread growth retardation and male hypogonadism. The next day, Professor Darby and I went to nearby villages to assess if, indeed, growth-retarded adolescents were clinically recognizable. We were accompanied by an Egyptian male nurse who spoke both English and Arabic.

Because of the striking clinical similarities between Iranian and Egyptian dwarfs, I was able to recognize several subjects who looked like 8- or 10-year-old boys, but whose chronological ages, on questioning, appeared to be 18 to 20 years. This assured me that, indeed, growth retardation and male hypogonadism were also prevalent in Egyptian villages. Following this experience, Professor Darby approved plans to investigate zinc metabolism in growth-retarded subjects.

The clinical features of Egyptian growth-retarded subjects were remarkably similar to those of the Iranians, except that the Iranian dwarfs had more pronounced hepatosplenomegaly, a history of geophagia, and no hookworm infection. The Egyptian subjects, by contrast, had both schistosomiasis and hookworm infestations but no history of geophagia.

A detailed investigation of the Egyptian cases was carried out with associates A. Miale, Z. Farid, H. H. Sandstead, and A. Schulert. The dietary history of the Egyptian subjects was similar to that of the Iranians. The consumption of animal protein was negligible, with the diet consisting mainly of bread and beans (*Vicia faba*).

Zinc concentrations in plasma, hair, and red cells were decreased, and ^{65}Zn studies showed that the plasma zinc turnover was greater, the 24 h exchange-

able pool was smaller, and the excretion of ^{65}Zn in stool and urine was less in the growth-retarded subjects in comparison to controls (Prasad, Miale, and Farid 1963). These studies established for the first time that zinc deficiency occurs in humans, without advanced cirrhosis. Liver function tests and biopsies revealed no evidence of cirrhosis in our subjects (Prasad et al. 1963b). Furthermore, in contrast to cirrhosis patients, who excreted abnormally high quantities of zinc in the urine, our patients excreted less zinc in urine than did control subjects. We also ruled out other chronic diseases that might have affected the serum zinc concentrations.

Serum iron was decreased, unsaturated iron binding capacity was increased, serum copper was slightly increased, and serum magnesium was normal in our subjects. Hair analysis for manganese, cobalt, molybdenum, and other elements revealed no difference when compared to the normal subjects. We found no evidence for deficiency of serum B_{12} , ascorbic acid, vitamin A, carotene, or folic acid.

Iranian physicians had commonly believed that growth retardation and sexual hypofunction were the results of visceral leishmaniasis and geophagia. We, however, found no evidence of leishmaniasis in Iran. The role of geophagia was unclear, although we suspected that excess phosphate in the clay prevented absorption of both iron and zinc. The predominantly wheat diet in the Middle East, now known to contain high quantities of phytate and fiber, most probably reduced the availability of zinc.

In Egypt the cause of dwarfism was commonly considered to be schistosomiasis, and in China investigators had also implicated schistosomiasis as a causative factor for growth retardation. Yet, because the Iranian subjects exhibited dwarfism but did not have schistosomiasis or hookworm infections, the question arose as to whether schistosomiasis was the fundamental cause of dwarfism in Egypt. We initiated an investigation to find the answer (Prasad, Schulert, and Miale 1963; Prasad 1966).

It was known that schistosomiasis or hookworm infection was nonexistent in the oasis villages of Kharga, located 500 kilometers (km) southwest of Cairo, although the people of Kharga are culturally and nutritionally similar to those in the delta region. Hence, we conducted a field study in Kharga on 16 patients with hypogonadism and dwarfism. Their anemia was mild, and none had either schistosomiasis or hookworm disease. Concentrations of iron and zinc in the serum were subnormal.

Because red blood cells are rich in both iron and zinc, blood loss due to hookworm and schistosomiasis in the delta villages contributed significantly to both iron and zinc deficiencies. But in Kharga, parasitic infections were not present, and the artesian spring, the principal source of water for the Kharga villages, revealed iron and zinc concentrations of 3,170 and 18 micrograms per liter ($\mu\text{g/l}$), respectively.

In Cairo, by contrast, the iron and zinc concentrations of drinking water were 70 and 400 µg/l, respectively. Thus, although the foods consumed by the subjects in both the delta region and the oasis villages were similar, those in the latter probably derived a significant amount of iron but no zinc from their water source. Consequently, a better iron status for individuals in Kharga villages, in comparison to those in delta villages, was due to higher iron intake from water and lack of blood loss due to parasites.

Although dwarfism and hypogonadism had previously been attributed to schistosomiasis in Egypt and China, our demonstration of the existence of such patients in Iran and Kharga, where schistosomiasis was absent, showed that this parasitic infection was not necessarily responsible for these clinical features.

We were, however, unable to account for the hepatosplenomegaly on the basis of liver disease. This left three possibilities: anemia, zinc deficiency, or a combination of the two. In each case, the size of the liver and spleen decreased significantly after zinc supplementation, suggesting that zinc deficiency may have played an as-yet undefined role in hepatosplenomegaly.

In the Middle East, we examined only male subjects, as females refused to participate. But later studies from Iran by J. A. Halsted and co-workers (1972) demonstrated that zinc deficiency was probably prevalent in females manifesting growth retardation.

Our further investigations in Egypt showed that the rate of growth was greater in patients who received supplemental zinc as compared with those given iron, or those with only an adequate animal protein diet (Prasad 1966; Sandstead et al. 1967). Pubic hair appeared in all subjects within 7 to 12 weeks after zinc supplementation was initiated. Genital size became normal, and secondary sexual characteristics developed within 12 to 24 weeks in all zinc supplemented patients. No such changes were observed in a comparable length of time in the iron-supplemented group or in the group on an animal protein diet alone. Thus, our studies demonstrated that both growth retardation and gonadal hypofunction in these subjects were related to zinc deficiency. The anemia was due to iron deficiency and responded to oral iron treatment.

Chronology of Other Observations

By using the dithizone technique, R. E. Lutz (1926) assayed zinc in various tissues and concluded that the body of a 70 kg man contained 2.2 grams (g) of zinc, a figure remarkably close to that which is accepted today. R. A. McCance and E. M. Widdowson (1942) were the first to report on the absorption and excretion of zinc, and showed that the principal route of zinc excretion was in the feces, with only a small amount lost in urine.

I. Vikbladh (1950) measured serum zinc concentration by dithizone technique and reported that the

level was 19.7 ± 0.24 micromoles per liter (µmol/l), a value in general agreement with those reported by using modern methods. Vikbladh (1951) also observed that the serum zinc concentration was decreased in many chronic diseases, including liver disease. B. L. Vallee and colleagues (1956) reported that the serum zinc concentration decreased in patients with cirrhosis and suggested that the hypozincemia of these subjects was conditioned by hyperzincuria.

Our studies in the early 1960s demonstrated for the first time the effects of zinc deficiency on human growth and gonadal development (Prasad, Miale, Farid et al. 1963a, 1963b; Prasad, Schultert, Miale et al. 1963; Sandstead et al. 1967), and that this deficiency may have various causes in different populations. It is now evident that nutritional, as well as conditioned, zinc deficiency may complicate many disease states in human subjects.

In 1968, R. A. MacMahon, M. L. Parker, and M. McKinnon (1968) first observed zinc deficiency in a patient who had steatorrhea. Subsequently, zinc deficiency was discovered to be common in patients with malabsorption syndromes (McClain, Adams, and Shedlofsky 1988).

V. Caggiano and co-workers (1969) were the first to report a case of zinc deficiency in the United States. The patient was Puerto Rican with dwarfism, hypogonadism, hypogammaglobulinemia, giardiasis, stronglyloidiasis, and schistosomiasis. Zinc supplementation resulted in improved growth and development.

In 1972, a number of Denver children from middle-class families were reported to exhibit evidence of symptomatic nutritional zinc deficiency (Hambidge et al. 1972). They consumed a predominantly cereal protein diet low in available zinc. Growth retardation, poor appetite, and impaired taste acuity were all related to a deficiency of zinc in the children, and they were corrected with supplementation. Symptomatic zinc deficiency in United States infants was also reported later by K. M. Hambidge, C. E. Casey, and N. J. Krebs (1983). In addition, our own recent studies in the United States have shown that zinc deficiency in the well-to-do elderly may be fairly prevalent (Prasad et al. 1993). Clearly, then, a substantial portion of the U.S. population may be at risk of zinc deficiency.

Meanwhile, Halsted and colleagues (1972) published a study involving a group of 15 men who were rejected at the Iranian army induction center because of "malnutrition." Two women were also included in their study. All were 19 or 20 years old, with clinical features similar to those reported earlier by A. S. Prasad and colleagues (Prasad et al. 1961, 1963a, 1963b). They were studied for 6 to 12 months. One group was given a well-balanced diet, containing ample animal protein plus a placebo capsule. A second group was given the same diet, plus a capsule of zinc sulfate containing 27 mg zinc. A third group received the diet for a year, with zinc supplementation during the last six months.

The zinc-supplemented subjects grew considerably faster and showed evidence of earlier onset of sexual function (as defined by nocturnal emission in males and menarche in females) than those receiving the well-balanced diet alone (Halsted et al. 1972).

A clinical picture, similar to those reported by our studies involving zinc-deficient dwarfs, has been observed in many developing countries. Therefore, it must be the case that various levels of zinc deficiency prevail in countries where diets depend too heavily on cereals.

P. M. Barnes and E. J. Moynahan (1973) studied a 2-year-old girl with severe acrodermatitis enteropathica, who was receiving diiodohydroxyquinoline and a lactose-deficient synthetic diet. The response to this therapy was not satisfactory. It was noted that the concentration of zinc in the patient's serum was profoundly decreased, and oral zinc sulfate was administered. The skin lesions and gastrointestinal symptoms cleared completely, and the girl was discharged from the hospital. When zinc was inadvertently omitted from the child's regimen, she suffered a relapse; however, she promptly responded to oral zinc again.

In the original report, the authors attributed the girl's zinc deficiency to the synthetic diet, but it soon became clear that zinc was fundamental in the pathogenesis of acrodermatitis enteropathica - a rare inherited disorder - and the clinical improvement reflected improvement in zinc status. This interesting observation was quickly confirmed in other patients throughout the world. The underlying pathogenesis of zinc deficiency in these patients is most likely dietary malabsorption of zinc, the mechanism of which remains to be determined.

R. G. Kay and C. Tasman-Jones (1975) reported the occurrence of severe zinc deficiency in subjects receiving total parenteral nutrition for prolonged periods without zinc. T. Arakawa, T. Tamura, and Y. Igarashi (1976) and A. Okada and co-workers (1976) announced similar findings in this circumstance. These observations have been documented by several investigators, and in the United States, zinc is routinely included in total parenteral fluids for subjects who are likely to receive such therapy for extended periods.

W. G. Klingberg, Prasad, and D. Oberleas (1976) were the first to report severe parakeratosis, alopecia, and retardation of growth and gonadal development in an adolescent with Wilson's disease who received penicillamine therapy. Zinc supplementation completely reversed these clinical manifestations.

Recent literature suggests that several findings in patients with sickle cell anemia, such as growth retardation, male hypogonadism, abnormal dark adaptation, and abnormal cell-mediated immunity, are related to a deficiency of zinc (Prasad et al. 1975, 1981; Warth et al. 1981; Prasad and Cossack 1984; Prasad et al. 1988). Hyperzincuria due to renal tubular dysfunction has been noted in such subjects, and this may be a

contributing factor in the pathogenesis of zinc deficiency. Hypogeusia, decreased serum testosterone level, and hyperprolactinemia due to zinc deficiency have been observed in male patients with chronic renal disease (Mahajan et al. 1979, 1980, 1982, 1985). Zinc supplementation has corrected the abnormalities that have been associated with these disparate circumstances.

During the past three decades, a spectrum of clinical deficiency of zinc in human subjects has been recognized. If the deficiency is severe, it may be life-threatening. The symptoms developed by severely zinc-deficient subjects include bullous-pustular dermatitis, diarrhea, alopecia, mental disturbances, and intercurrent infections due to cell-mediated immune disorders. These manifestations are seen in patients with acrodermatitis enteropathica, following total parenteral nutrition (without zinc), and after penicillamine therapy.

Growth retardation, male hypogonadism, skin changes, poor appetite, mental lethargy, abnormal adaptation to darkness, and delayed wound healing are some of the indicators of moderate zinc deficiency in human subjects. Causes of moderate zinc deficiency that have been well documented include nutritional factors, malabsorption, sickle cell disease, chronic renal disease, and other debilitating conditions.

The beneficial effect of zinc in healing wounds of patients with zinc deficiency was first reported by W. J. Pories and W. H. Strain (1966). The symptom of abnormalities of taste was first related to a deficiency of zinc in humans by R. I. Henkin and D. F. Bradley (1969), and such abnormalities, which are reversible by zinc supplementation, have been observed in patients with chronic renal disease (Mahajan et al. 1980).

Marginal Deficiency of Zinc

Although the importance of zinc to human health has now been elucidated and its deficiency recognized in several clinical conditions, it was only recently that an experimental human model was established to permit a study of the specific effects of a mild zinc deficiency (Prasad, Rabbani, and Abbasi 1978; Abbasi et al. 1980; Rabbani et al. 1987; Prasad et al. 1988).

We did this by developing a semisynthetic soy-protein-based diet that supplies 3 to 5 milligrams of zinc per day (mg zinc/d) (Rabbani et al. 1987). All other nutrients in the diet are consistent with the RDA (1974, 1989). Male volunteers, ages 20 to 45 years, were first given a hospital diet containing animal protein for 4 to 8 weeks, which averaged 12 mg zinc/d. After that, the subjects received the experimental diet containing 3 to 5 mg zinc/d, which continued for 28 weeks. Following this period, the volunteers received a daily 27 mg zinc supplement for 12 weeks while still consuming the experimental diet. Throughout the study, all nutrients, including protein, amino acids,

vitamins, and minerals (both macro- and microelements), were kept constant, except zinc, which was varied as outlined above. By this technique, we were able to induce a specific zinc deficiency in men.

Our dietary manipulation created a negative zinc balance of approximately 1 mg per day, and we calculated that in a six-month period a total of about 180 mg of negative zinc balance was achieved. A 70 kg adult male contains approximately 2,300 mg of zinc, and, therefore, a loss of 180 mg of zinc would seem to be only 8 percent of the total body zinc. But this is not necessarily the case. Approximately 28 percent of the zinc in the human body resides in bone, 62 percent in muscle, 1.8 percent in the liver, and 0.1 percent in the plasma pool. Only 10 percent of the total body zinc pool exchanges with an isotopic dose within a week's time (Prasad et al. 1963a; Foster et al. 1979).

In an adult animal model, zinc concentrations in muscle and bone do not change as a result of mild or marginal zinc deficiency. In cases of mild or marginal zinc deficiency, one cannot expect a uniform distribution of the deficit over the entire body pool, and most likely the compartments with high turnover rates (liver and peripheral blood cells, such as lymphocytes, granulocytes, and platelets) suffer a disproportionate deficit. Thus, if one were to consider that only 200 to 400 mg zinc, which is represented by liver zinc and the mobile exchangeable pool, is the critical pool, a negative balance of 180 mg from this pool may be a considerable fraction.

Our studies in this model have indicated that a mild or marginal deficiency of zinc in humans is characterized by neurosensory changes, oligospermia, decreased serum testosterone concentration, hyperammonemia, decreased lean body mass, decreased serum thymulin activity, decreased IL-2 production by peripheral blood mononuclear cells, decreased NK cell activity, and alterations in T-cell subpopulations. All of these manifestations can be corrected by zinc supplementation.

When zinc deficiency was very mild (5.0 mg zinc intake during the 20- to 24-week zinc-restricted period), the plasma zinc concentration remained more or less within the normal range, whereas the zinc concentration of lymphocytes and granulocytes declined (Meftah et al. 1991). Within 8 weeks of zinc restriction, the activity of lymphocyte ecto 5' nucleotidase (5'NT), serum thymulin activity, and IL-2 production by peripheral blood mononuclear cells decreased, and the intestinal absorption of ^{70}Zn increased significantly, suggesting that lymphocytes, thymus, and intestinal cells are very sensitive to zinc restriction (Meftah et al. 1991; Lee et al. 1993; Prasad et al. unpublished observation).

Biochemical Advances in Zinc Metabolism

D. Keilin and J. Mann (1940) were the first to demonstrate that carbonic anhydrase was a zinc metalloen-

zyme. Over the next 20 years, only five additional zinc metalloenzymes were identified, but in the last 30 years the total number has greatly increased. If related enzymes for different species are included, more than 200 zinc metalloenzymes are now known to exist (Chesters 1982; Galdes and Vallee 1983).

I. Lieberman and co-workers (1963) have shown that several enzymes necessary for nucleic acid synthesis in microorganisms require zinc. It is now well known that zinc is needed for DNA polymerase 1 (in *Escherichia coli*), bacterial RNAPolymerase (in *E. coli*), and reverse transcriptase (in avian myeloblastosis virus) (Wu and Wu 1983).

Until 1965, there was no evidence that zinc-dependent enzymes were adversely affected as a result of zinc deficiency. Our investigations then demonstrated that the activity of various zinc-dependent enzymes was reduced in the testes, bones, esophagus, and kidneys of zinc-deficient rats in contrast to their pair-fed controls, and that this reduction of activity correlated with the decreased zinc content of the tissues (Prasad, Oberleas, Wolf et al. 1967).

Several studies have shown that zinc deficiency in animals impairs the incorporation of labeled thymidine into DNA. This effect has been detected within a few days of the institution of a zinc deficient diet in experimental animals, suggesting that dietary zinc deficiency may result in an immediate impairment of DNA synthesis. Prasad and Oberleas (1974) provided evidence that this early reduction in DNA synthesis was due to an adverse effect of zinc restriction on the activity of deoxythymidine kinase. These results were confirmed by I. E. Dreosti and L. S. Hurley (1975), who showed that the activity of deoxythymidine kinase in 12-day-old fetuses taken from females exposed to a dietary zinc deficiency during pregnancy was significantly lower than in ad-libitum-fed and restricted-fed controls.

Zinc and Immunity

P. J. Fraker, S. Hass, and R. W. Luecke (1977) revealed that severely and marginally zinc-deficient young adult-A/Jax mice have abnormal T-helper cell function. In addition, it is now known that other T-lymphocyte-mediated functions are found to be adversely affected by zinc deficiency. By using the young adult mouse as a model, it was demonstrated that a moderate period of suboptimal zinc administration causes thymic atrophy, lymphopenia, and alterations in the proportions of the various subsets of lymphocytes and mononuclear phagocytes (Fraker et al. 1986). As a result, antibody-mediated responses to both T-cell-dependent and T-cell-independent antigens are significantly reduced. Cytolytic T-cell responses, NK-cell activity, and delayed-type-hypersensitivity (DTH) reactions are also depressed.

In humans, patients with acrodermatitis enteropathica (a genetic disorder of zinc absorption) exhibit

atrophic thymus, lymphopenia, anergic DTH responses, and reduced NK-cell activity (Fraker et al. 1986). Impaired DTH responses, correctable with zinc supplementation, were reported in zinc-deficient sickle cell anemia patients (Ballester and Prasad 1983), as were decreased NK-cell activity, decreased IL-2 activity, decreased serum thymulin activity, and alterations in lymphocyte subpopulations (Prasad et al. 1988).

Metallothionein

Metallothionein (MT) was discovered in 1957. M. Margoshes and Vallee (1957) identified a cadmium-binding protein in equine kidney cortex responsible for the natural accumulation of cadmium in the tissues. Metal and sulfur content are extremely high in MTs. In human cells, expression of the ISO-MT genes appears to be regulated differentially by cadmium, zinc, and glucocorticoids, and ISO-MT genes are indications for tissue-specific expression (Kagi and Schaffer 1988). A number of studies have led to the identification of various DNA segments serving as promoter sites in the 5' region of various MT genes in induction by metal ions and hormones. In the mouse MT-1 gene, the functional metal responsive promoter is composed of a set of four closely related metal-regulatory elements, each made up of eight nucleotides and localized near the TATA box.

Zinc may be the regulator of the mRNA strands responsible for de novo synthesis of MT in intestinal cells (Cousins 1979). It has been suggested that MT programs the fluctuating levels of zinc in and out of intestinal cells and plays an important role in regulating the absorption and/or excretion of not only zinc but also cadmium and copper.

Zinc and Gene Expression

The importance of zinc in DNA-binding proteins as regulators of gene expression has been recently recognized (Brown, Sander, and Argos 1985; Miller, Mclachlan, and Klug 1985; Klug and Rhodes 1987). The first zinc-finger protein to be recognized was transcription factor-III α of *Xenopus laevis*, which contained tandem repeats of segments with 30 amino acid residues, including pairs of histidines and cysteines (Miller et al. 1985). The presence of zinc in these proteins is essential for site-specific binding to DNA and gene expression. The zinc ion apparently serves as a strut that stabilizes folding of the domain into a finger-loop, which is then capable of site-specific binding to double-stranded DNA. The zinc-finger loop proteins provide one of the fundamental mechanisms for regulating gene expression of many proteins. In humans, the steroid hormones (and related compounds, such as thyroid hormones, cholecalciferol, and retinoic acid) enter cells by facilitated diffusion and combine with respective receptors (which contain the DNA-binding domain of the zinc-finger loops) either before or after entering the nucleus. Complexing of a hormone by its

specific receptor evidently initiates a conformation change that exposes the zinc-finger loops, so that they bind to high-affinity sites on DNA and regulate gene expression (Hollenberg et al. 1985; Hughes et al. 1988; Sunderman and Barber 1988).

Interaction of Zinc with Other Elements

Zinc blocks the absorption of dietary copper, and also copper in the endogenous secretions (Brewer et al. 1988). Earlier, Prasad, Brewer, Schoomaker et al. (1978) observed that when subjects with sickle cell anemia were treated with 150 mg zinc/d in divided doses in order to reduce the number of irreversible sickle cells in the peripheral blood, they showed a decrease in the concentration of serum copper and ceruloplasmin. This observation led us to consider treatment of Wilson's disease patients with zinc. Our studies showed that zinc therapy in Wilson's disease patients leads to a negative copper balance, most likely by induction of MT synthesis in the intestines, whereby copper is sequestered and ultimately excreted in the feces (Brewer et al. 1983, 1987). According to our experience, zinc is an effective copper removal agent, is well tolerated, and prevents accumulation of copper in the liver (Brewer et al. 1983, 1987, 1988).

Ananda S. Prasad

This work was supported in part by grants from NIH/NIDDK (No. DK31401); NIH/NCI (No. CA 43838); the Food and Drug Administration (No. FDA-U-000457); Labcatal Laboratories, Paris, France; and the Veterans Administration Medical Research Service.

Bibliography

- Abbasi, A. A., A. S. Prasad, P. Rabbani, and E. DuMouchelle. 1980. Experimental zinc deficiency in man: Effect on testicular function. *Journal of Laboratory and Clinical Medicine* 96: 544-50.
- Arakawa, T., T. Tamura, and Y. Igarashi. 1976. Zinc deficiency in two infants during parenteral alimentation for diarrhea. *American Journal of Clinical Nutrition* 29: 197-204.
- Ballester, O. F., and A. S. Prasad. 1983. Anergy, zinc deficiency and decreased nucleoside phosphorylase activity in patients with sickle cell anemia. *Annals of Internal Medicine* 98: 180-2.
- Barnes, P. M., and E. J. Moynahan. 1973. Zinc deficiency in acrodermatitis enteropathica: Multiple dietary intolerance treated with synthetic zinc. *Proceedings of the Royal Society of Medicine* 66: 327-9.
- Brewer, G. J., G. M. Hill, R. D. Dick, et al. 1987. Treatment of Wilson's disease with zinc. III. Prevention of reaccumulation of hepatic copper. *Journal of Laboratory and Clinical Medicine* 109: 526-31.
- Brewer, G. J., G. M. Hill, A. S. Prasad, et al. 1983. Oral zinc therapy for Wilson's disease. *Annals of Internal Medicine* 99: 314-20.
- Brewer, G. J., V. A. Yuzbasiyan, V. Iyengar, et al. 1988. Regulation of copper balance and its failure in humans. In *Essential and toxic trace elements in human health and disease*, ed. A. S. Prasad, 95-103. New York.

- Brown, R. S., C. Sander, and P. Argos. 1985. The primary structure of transcription factor TF-III A has 12 consecutive repeats. *FEBS Letter* 186: 271-4.
- Caggiano, V., R. Schnitzler, W. Strauss, et al. 1969. Zinc deficiency in a patient with retarded growth, hypogonadism, hypogammaglobulinemia, and chronic infection. *American Journal of Medical Science* 257: 305-19.
- Chesters, J. K. 1982. Metabolism and biochemistry of zinc. In *Clinical, biochemical and nutritional aspects of trace elements*, ed. A. S. Prasad, 221-38. New York.
- Cousins, R. J. 1979. Regulatory aspects of zinc metabolism in liver and intestine. *Nutrition Review* 34: 97-103.
- Dreosti, I. E., and L. S. Hurley. 1975. Depressed thymidine kinase activity in zinc deficient rat embryos. *Proceedings of the Society of Experimental and Biological Medicine* 150: 161-5.
- Foster, D. M., R. L. Aamodt, R. I. Henkin, and M. Berman. 1979. Zinc metabolism in humans. A kinetic model. *American Journal of Physiology* 237: R340-9.
- Fraker, P. J., M. E. Gershwin, R. A. Good, and A. S. Prasad. 1986. Interrelationship between zinc and immune function. *Federation Proceedings*, 45: 1474-9.
- Fraker, P. J., S. Hass, and R. W. Luecke. 1977. Effect of zinc deficiency on the immune response of the young adult A/Jax mouse. *Journal of Nutrition* 107: 1889-95.
- Galdes, A., and B. L. Vallee. 1983. Categories of zinc metalloenzymes. In *Metal ions in biological systems*, ed. H. Sigel, 1-54. New York.
- Halsted, J. A., H. A. Ronaghy, P. Abadi, et al. 1972. Zinc deficiency in man: The Shiraz experiment. *American Journal of Medicine* 53: 277-84.
- Hambidge, K. M., C. E. Casey, and N. J. Krebs. 1983. Zinc. In *Trace elements in human and animal nutrition*, Vol 2, ed. M. Mertz, 1-137. New York.
- Hambidge, K. M., C. Hambidge, M. Jacobs, and J. D. Baum. 1972. Low levels of zinc in hair, anorexia, poor growth, and hypogeusia in children. *Pediatric Research* 6: 868-74.
- Henkin, R. I., and D. F. Bradley. 1969. Regulation of taste acuity by thiols and metal ions. *Proceedings of the National Academy of Science USA* 62: 30-7.
- Hollenberg, S. M., C. Weinberger, E. S. Ong, et al. 1985. Primary structure and expression of a functional human glucocorticoid receptor cDNA. *Nature* 318: 635-41.
- Hughes, M. R., P. J. Malloy, D. G. Kieback, et al. 1988. Point mutations in the human vitamin D receptor gene associated with hypocalcemia rickets. *Science* 242: 1702-5.
- Kagi, J. H. R., and A. Schaffer. 1988. Biochemistry of metallothionein. *Biochemistry* 27: 8509-15.
- Kay, R. G., and C. Tasman-Jones. 1975. Zinc deficiency and intravenous feeding. *Lancet* 2: 605-6.
- Keilin, D., and J. Mann. 1940. Carbonic anhydrase. Purification and nature of the enzyme. *Biochemistry Journal* 34: 1163-76.
- Klingberg, W. G., A. S. Prasad, and D. Oberleas. 1976. Zinc deficiency following penicillamine therapy. In *Trace elements in human health and disease*, Vol. 1, ed. A. S. Prasad, 51-65. New York.
- Klug, A., and A. Rhodes. 1987. "Zinc-fingers." A novel protein motif for nucleic acid recognition. *Trends in Biochemical Science* 12: 461-9.
- Lee, D.-Y., A. S. Prasad, H. Constance, et al. 1993. Homeostasis of zinc in marginal human zinc deficiency: Role of absorption and endogenous excretion of zinc. *Journal of Laboratory and Clinical Medicine*. 122: 549-56.
- Lemann, I. I. 1910. A study of the type of infantilism in hookworm disease. *Archives of Internal Medicine* 6: 139-46.
- Lieberman, I., R. Abrams, N. Hunt, and P. Ove. 1963. Levels of enzyme activity and deoxyribonucleic acid synthesis in mammalian cells cultured from the animal. *Journal of Biology and Chemistry* 238: 3955-62.
- Lutz, R. E. 1926. The normal occurrence of zinc in biological materials: A review of the literature, and a study of the normal distribution of zinc in the rat, cat and man. *Biochemistry Journal* 34: 991-7.
- MacMahon, R. A., M. L. Parker, and M. McKinnon. 1968. Zinc treatment in malabsorption. *Medical Journal of Australia* 2: 210-12.
- Mahajan, S. K., R. J. Hamburger, W. Flamenbaum, et al. 1985. Effect of zinc supplementation on hyperprolactinemia in uremic men. *Lancet* 2: 750-1.
- Mahajan, S. K., A. S. Prasad, W. A. Briggs, and F. D. McDonald. 1982. Correction of taste abnormalities and sexual dysfunction by zinc (Zn) in uremia. A double-blind study. *Annals of Internal Medicine* 97: 357-71.
- Mahajan, S. K., A. S. Prasad, J. Lambujon, et al. 1980. Improvement of uremic hypogeusia by zinc: A double-blind study. *American Journal of Clinical Nutrition* 33: 1517-21.
- Mahajan, S. K., A. S. Prasad, P. Rabbani, et al. 1979. Zinc metabolism in uremia. *Journal of Laboratory and Clinical Medicine* 94: 693-8.
- Margoshes, M., and B. L. Vallee. 1957. A cadmium protein from equine kidney cortex. *American Journal of the Chemical Society* 79: 4813-14.
- McCance, R. A., and E. M. Widdowson. 1942. The absorption and excretion of zinc. *Biochemistry Journal* 36: 692-6.
- McClain, C. J., L. Adams, and S. Sheddlofsky. 1988. Zinc and the gastrointestinal system. In *Essential and toxic trace elements in human health and disease*, ed. A. S. Prasad, 55-73. New York.
- Meftah, S., A. S. Prasad, D.-Y. Lee, and G. J. Brewer. 1991. Ecto 5' nucleotidase (5'NT) as a sensitive indicator of human zinc deficiency. *Journal of Laboratory and Clinical Medicine* 118: 309-16.
- Miller, J., A. D. McLachlan, and A. Klug. 1985. Repetitive zinc-binding domains in the protein transcription factor III-A for xenopus oocytes. *EMBO Journal* 4: 1609-14.
- Minnich, V., A. Okevgola, Y. Tarcon, et al. 1968. The effect of clay on iron absorption as a possible cause for anemia in Turkish subjects with pica. *American Journal of Clinical Nutrition* 21: 78-86.
- National Research Council. 1974. *Recommended dietary allowances*. Eighth edition. Washington, D.C.
1989. *Recommended dietary allowances*. Tenth edition. Washington, D.C.
- O'Dell, B. L., P. O. Newberne, and J. E. Savage. 1958. Significance of dietary zinc for the growing chicken. *Journal of Nutrition* 65: 503-18.
- O'Dell, B. L., and J. E. Savage. 1960. Effect of phytic acid on zinc availability. *Proceedings of the Society of Experimental and Biological Medicine* 103: 304-6.
- Okada, A., Y. Takagi, T. Itakura, et al. 1976. Skin lesions during intravenous hyperalimentation: Zinc deficiency. *Surgery* 80: 629-35.
- Pories, W. J., and W. H. Strain. 1966. Zinc and wound healing. In *Zinc metabolism*, ed. A. S. Prasad, 378-94. Springfield, Ill.
- Prasad, A. S. 1966. Metabolism of zinc and its deficiency in human subjects. In *Zinc metabolism*, ed. A. S. Prasad, 250-302. Springfield, Ill.
1988. Clinical spectrum and diagnostic aspects of human zinc deficiency. In *Essential and toxic trace elements in human health and disease*, ed. A. S. Prasad, 3-53. New York.

- Prasad, A. S., A. A. Abbasi, P. Rabbani, and E. DuMouchelle. 1981. Effect of zinc supplementation on serum testosterone level in adult male sickle cell anemia subjects. *American Journal of Hematology* 19: 119-27.
- Prasad, A. S., G. J. Brewer, E. B. Schoomaker, and P. Rabbani. 1978. Hypocupremia induced by zinc therapy in adults. *Journal of the American Medical Association* 240: 2166-8.
- Prasad, A. S., and Z. T. Cossack. 1984. Zinc supplementation and growth in sickle cell disease. *Annals of Internal Medicine* 100: 367-71.
- Prasad, A. S., J. T. Fitzgerald, J. W. Hess, et al. 1993. Zinc deficiency in the elderly. *Nutrition* 9: 218-24.
- Prasad, A. S., J. A. Halsted, and M. Nadimi. 1961. Syndrome of iron deficiency anemia, hepatosplenomegaly, hypogonadism, dwarfism and geophagia. *American Journal of Medicine* 31: 532-46.
- Prasad, A. S., S. Meftah, J. Abdallah, et al. 1988. Serum thymulin in human zinc deficiency. *Journal of Clinical Investigation* 82: 1202-10.
- Prasad, A. S., A. Miale, Jr., Z. Farid, et al. 1963. Zinc metabolism in patients with the syndrome of iron deficiency anemia, hypogonadism, and dwarfism. *Journal of Laboratory and Clinical Medicine* 61: 537-49.
- 1963b. Biochemical studies on dwarfism, hypogonadism, and anemia. *Archives of Internal Medicine* 111: 407-28.
- Prasad, A. S., and D. Oberleas. 1974. Thymidine kinase activity and incorporation of thymidine into DNA in zinc-deficient tissue. *Journal of Laboratory and Clinical Medicine* 83: 634-9.
- Prasad, A. S., D. Oberleas, P. Wolf, et al. 1967. Studies on zinc deficiency: Changes in trace elements and enzyme activities in tissues of zinc-deficient rats. *Journal of Clinical Investigation* 46: 549-57.
- Prasad, A. S., P. Rabbani, A. Abbasi, et al. 1978. Experimental zinc deficiency in humans. *Annals of Internal Medicine* 89: 483-90.
- Prasad, A. S., E. B. Schoomaker, J. Ortega, et al. 1975. Zinc deficiency in sickle cell disease. *Clinical Chemistry* 21: 582-7.
- Prasad, A. S., A. R. Schulert, A. Miale Jr., et al. 1963. Zinc and iron deficiencies in male subjects with dwarfism and hypogonadism but without acylostomiasis and schistosomiasis or severe anemia. *American Journal of Clinical Nutrition* 12: 437-44.
- Rabbani, P. I., A. S. Prasad, R. Tsai, et al. 1987. Dietary model for production of experimental zinc deficiency in man. *American Journal of Clinical Nutrition* 45: 1514-25.
- Raulin, J. 1869. Chemical studies on vegetation. *Annals of Science Nature* 11 (in French): 93-9.
- Reimann, F. 1955. Growth anomalies and malformations in iron-deficient states. *Proceedings of the 5th Kongr Eur Gesellschaft Haematology*, Freiburg, FRG: HM Keller, (in German) 546-50.
- Sandstead, H. H., A. S. Prasad, A. S. Schulert, et al. 1967. Human zinc deficiency, endocrine manifestations and response to treatment. *American Journal of Clinical Nutrition* 20: 422-42.
- Sommer, A. L., and C. B. Lipman. 1926. Evidence of indispensable nature of zinc and boron for higher green plants. *Plant Physiology* 1: 231-49.
- Sunderman, F. W., Jr., and A. M. Barber. 1988. Finger loops, oncogenes, and metals. *Annals of Clinical and Laboratory Science* 18: 267-88.
- Todd, W. R., C. A. Elvehjem, and E. B. Hart. 1934. Zinc in the nutrition of the rat. *American Journal of Physiology* 107: 146-56.
- Tucker, H. F., and W. D. Salmon. 1955. Parakeratosis or zinc deficiency disease in pigs. *Proceedings of the Society of Experimental Biological Medicine* 88: 613-16.
- Vallee, B. L., W. E. C. Wacker, A. F. Bartholomay, et al. 1956. Zinc metabolism in hepatic dysfunction. I. Serum zinc concentrations in Laennec's cirrhosis and their validation by sequential analysis. *New England Journal of Medicine* 255: 403-8.
- Vikbladh, I. 1950. Studies on zinc in blood. *Scandinavian Journal of Clinical Laboratory Investigation* 2: 143-8.
1951. Studies on zinc in blood. *Scandinavian Journal of Clinical Laboratory Investigation* (Suppl) 3: 1-74.
- Warth, J. A., A. S. Prasad, F. Zwas, et al. 1981. Abnormal dark adaptation in sickle cell anemia. *Journal of Laboratory and Clinical Medicine* 98: 189-94.
- Wu, F. Y., and C. Wu. 1983. The role of zinc in DNA and RNA polymerases. In *Metal ions in biological systems*, ed. H. Sigel, 157-92. New York.



IV.C

Proteins, Fats, and Essential Fatty Acids

IV.C.1 Essential Fatty Acids

The history of the scientific documentation of the need for fat in the diet began with the early nineteenth-century work of Michel Eugene Chevreul (Mayer and Hanson 1960). He showed that lard contained a solid fat, which he termed stearine, and a liquid fat he called elaine (later shown to be the isomer of oleine), and in 1823, this work was published in a treatise, *Chemical Investigations of Fats of Animal Origin*. Chevreul also crystallized potassium stearate, naming it “mother-of-pearl” and calling its acidified product “margarine” (from the Greek word for mother-of-pearl). In addition, Chevreul isolated various acids from fats and distinguished them on the basis of their melting points.

Meanwhile in 1822, Edmund Davy had reported that iodine would react with fats, and by the end of the century, work by L. H. Mills and Baron Hubl led to the procedure devised by J. J. A. Wijs in 1898 for determining a fat’s “iodine value” or “iodine number” – a measure of the extent to which a fat is unsaturated, based on its uptake of iodine. Highly saturated coconut oil, for example, has an iodine number of 8 to 10, whereas that of highly unsaturated linseed oil ranges from 170 to 202.

Phospholipids were described in 1846 by N. T. Goble, who found that egg yolk had a substance that contained nitrogen and phosphorus in addition to glycerol and fatty acids. He named it lecithin. The nitrogenous base was shown to be choline by A. Strecker in 1868, and J. W. L. Thudichem described kephalin in 1884 (Mayer and Hanson 1960).

These early advances in chemistry and methodology were necessary before the major leap to studying the fats in nutrition could be taken. Indeed, advances in chemistry and technology have preceded key discoveries in the study of essential fatty acids and their functions throughout its history.

The Concept of Essential Fatty Acids

During the late 1920s, several groups of investigators explored questions of the nutritional value of fat that went beyond the knowledge that dietary fats provided energy and contained vitamins A and D. Among them were Herbert M. Evans and George O. Burr, who experimented with diets that were sufficiently purified to exclude fat. In three published papers (1927a; 1927b; 1928), they described a previously unknown deficiency disease that resulted from an absence of dietary fats and suggested the existence of a new vitamin, “vitamine F” Shortly after this, at Yale University, Ava Josephine McAmis, William E. Anderson, and Lafayette B. Mendel determined that a nonsaponifiable fraction of cod liver oil delivered vitamin A – but no fat – to experimental rats, and that slightly better growth was achieved when about 20 milligrams (mg) of peanut oil was fed to the rats along with the cod liver oil fraction. Animals receiving the most peanut oil grew best. The researchers wrote, with exemplary caution, that “whether this apparent beneficial effect of a small amount of fat is due to its content of vitamin A or other vitamins, or to its acting as a vehicle for the fat-soluble vitamins, or whether fat per se is essential, is not conclusively demonstrated” (McAmis, Anderson, and Mendel 1929: 262).

At the same time, George and Mildred Burr, at the University of Minnesota, published their results from feeding a very low fat diet to rats (1929). They concluded that there was, indeed, a requirement for fat in the diet and also believed that they had discovered a new deficiency disease curable by the feeding of small amounts of unsaturated fats or pure “linolic [*sic*]” acid. The following year, they coined the term “essential fatty acids” (1930).

The symptoms observed in rats that were considered indicative of dietary insufficiency included late failure of growth, kidney lesions, abnormal water consumption (because of excessive extradermal water loss), scaly skin, and necrotic tails. However, the concept of the essentiality of fatty acids was not immedi-

ately accepted. Critics pointed out that the skin lesions of rats, for example, were also seen with some B-vitamin deficiencies and thus not specific to fat deficiency. In addition, there was confusion surrounding the nature of the "fat-free" diets fed to rats because those consuming cornstarch managed some growth, whereas those on sucrose did not.

These questions were explained by the studies of Evans and Samuel Lepkovsky (at the University of California, Berkeley), who, first of all, showed that there was enough fat in cornstarch to support some growth in rats (1932a). They determined further that, if saturated fat (in this case, coconut oil) rather than carbohydrate was fed to the rats, deficiency symptoms became evident even more rapidly (1932b). In still another experiment (1932c), they isolated the fatty acid methyl esters from carcasses of rats that had been on either the fat-free or the supplemented diets, fed the esters to other rats, and discovered that the iodine numbers of fatty acids from fat-free-fed animals were higher than those from rats fed supplemented diets, thus showing that the degree of unsaturation is not a criterion of fatty acid essentiality.

The fat from the rats fed fat-free diets contained more unsaturated fatty acids than that from other rats at weaning, but the unsaturated fatty acids present did not relieve the deficiency symptoms, which was the first indication of the biosynthesis of eicosatrienoic acid. This phenomenon was later observed by Raymond Reiser and colleagues (1951) and finally explained by A. J. Fulco and J. F. Mead (1959). In the meantime, Evans, Lepkovsky, and E. A. Murphy had added both failure of reproduction and lactation in female rats (1934a, 1934b) and sterility in male rats (1934c) to the list of deficiency symptoms.

Requirement for Essential Fatty Acids

The reports of skin lesions in rats fed fat-free diets caused pediatrician Arild E. Hansen (1933) to suspect that infants suffering from eczema had an unsatisfied fat requirement. Hansen treated the condition with various oils, and shortly thereafter, Theodore Cornbleet reported the highly successful treatment of eczema by dietary supplementation with corn oil (1935). He also noted that this supplementation brought relief to his patients with asthma.

Later, in collaboration with Hilda F. Wiese and others at the University of Texas School of Medicine in Galveston, Hansen studied essential fatty acid requirements using dogs as the experimental model (Hansen and Wiese 1951). As their methods for quantifying fatty acids progressed from fractionation of methyl esters to spectrophotometric analysis of the alkali-conjugated fatty acids (Wiese and Hansen 1953), Hansen and Wiese were able to determine serum levels of unsaturated fatty acids in both poorly

nourished (Hansen and Wiese 1954) and healthy (Wiese, Gibbs, and Hansen 1954) children. This work resulted in hard evidence of a deficiency of essential fatty acids that could occur in infants; it also resulted in recommendations for dietary intake of linoleic acid based upon the serum levels of di-, tri-, and tetraenoic acids of infants fed formulas containing different sources of fat (Wiese, Hansen, and Adams 1958). The researchers concluded that the dietary linoleate needed to provide optimum serum concentrations of polyunsaturated fatty acids was about 4 percent of the total calories. This advance in understanding the biochemistry of fatty acids in serum led to the use of biochemical criteria for defining nutrient deficiency.

Interrelationships among Fatty Acids

The alkaline isomerization method for analyzing specific fatty acids had led to the deduction that arachidonic (tetraenoic) acid was formed from dietary linoleic acid and that pentaenoic and hexaenoic acids were formed from dietary linolenic acid (Rieckehoff, Holman, and Burr 1949; Widmer and Holman 1950). Moreover, studies by Ralph Holman and his group at the Hormel Institute, University of Minnesota, also determined that linoleic and linolenic acids were not interconvertible. These findings were confirmed by the use of radioisotopically labeled fatty acids (Steinberg et al. 1956), and radioisotope tracer methods were used to define the source of Evans and Lepkovsky's (1932c) noncurative polyunsaturated fatty acid. A trienoic acid was identified by Fulco and Mead (1959) as 5,8,11-eicosatrienoic acid, which they found was derived from oleic acid.

During the 1960s, gas-liquid chromatography became the method of choice for the identification and quantification of fatty acids. A series of dose-response studies using single pure unsaturated fatty acids was conducted by Holman and his associates (Holman 1964), who developed interaction relationships which showed that the ratio of triene (eicosatrienoic acid) to tetraene (arachidonic acid) was proportional to the linoleate concentration in the diet of rats. They also determined that the diene and triene fatty acids were competitive with each other in conversions to longer-chain metabolites. For instance, 0.1 percent of energy as linolenate inhibited metabolism of linoleate by 50 percent, whereas it required 3.2 percent of linoleate to inhibit linolenate conversion by 50 percent. Interpretation of their many experiments led to the presentation of the sequence of chain elongation and desaturation of linoleic acid as 18:2 to 18:3 or 20:3; 18:3 to 20:3 to 20:4 or 22:3; 20:4 to 22:4 to 24:4 or 22:5. In the absence of linoleate and linolenate, the conversion of endogenous oleate (18:1) to 20:3 and 22:3 becomes dominant.

Nomenclature for Essential Fatty Acids

In studies prior to those of Holman and others in the 1960s, the original common names of the fatty acids were used. However, as the complexities of their double-bond configurations began to define their places in metabolism, it became necessary to establish simple and clear nomenclature. Diene, triene, monoene, and saturated, along with chain length, were no longer sufficient to describe physiologically important activities of the compounds. The locations of the double bonds were clarified, but the delta notation for the position of the double bonds was confusing when applied to chain elongation and desaturation. Holman (1964) used nomenclature based on the position of the double bond in relation to the *n*th (or omega) carbon. The linoleic acid family began with a double bond at the *n* minus 6 position; therefore, it was called the omega 6 family. The naturally occurring polyunsaturated fatty acids have methyl interrupted, rather than conjugated spacing; so linoleate is 18:2 omega 6,9. In the same pattern, linolenic acid is 18:3 omega 3,6,9. The importance of the specificity of desaturases was yet to be explained.

Importance of Human Requirements

Even though it had become accepted that linoleate was an essential nutrient, recommendations for dietary consumption were considered only for infants. Hansen, the pediatrician, agreed with Holman, the biochemist, that 0.5 to 1 percent of energy was enough, based upon keeping the triene-tetraene ratio below 0.4 (Holman 1973). By this time, the feeding of a fat-deficient diet to human subjects was not acceptable, so human studies were conducted using supplements; and, of course, diagnosis of malnourished infants also yielded findings.

The definitive proof of essential fatty acid requirement, however, came with the advent of intravenous feeding and its ability to provide total nutritional support (made possible by the method of implantation of a catheter in the superior vena cava, allowing infusion of hyperosmolar fluid) for long periods of time (Dudrick et al. 1968). Early formulas used glucose-protein hydrolysate fluid with electrolytes, minerals, and vitamins added, and reports of essential fatty acid deficiency symptoms in infants began to appear (Hallberg, Schuberth, and Wretling 1966). Fatty acid analyses were made when a case came to the attention of the Hormel Institute group, which reported that after 100 days of total parenteral nutrition, the infant had a triene-tetraene ratio of 18 and extreme scaliness of skin (Paulsrud et al. 1972).

The first efforts to add lipids to intravenous formulas were unsuccessful. An emulsion containing cottonseed oil proved unacceptable because of toxic reactions, and that experience delayed general use of lipid emulsions (Alexander and Ziene 1961). Indeed, as late as 1973, the U.S. Food and Drug Administration had

not approved the addition of fat preparations to parenteral formulas, and reports continued of essential fatty acid deficiency in infants (White et al. 1973). A 10 percent soybean oil emulsion, which employed a nontoxic emulsifying agent, egg phospholipid, and a smaller fat particle (0.5 microns in diameter), finally proved to be acceptable (Bivins et al. 1980).

Infants were not the only ones at risk for essential fatty acid deficiency from total parenteral nutrition. Adults with lesions of the gastrointestinal tract were reported to have biochemically defined linoleate deficiency (Wapnick, Norden, and Venturas 1974). Fourteen patients had triene-tetraene ratios with an average of greater than 2. Adults who had sufficient stores of adipose tissue fat were not believed to be at risk of essential fatty acid deficiency until the use of total parenteral nutrition (TPN) showed that this was not necessarily the case. When glucose and amino acids were infused continuously, lipolysis of adipose tissue fat was suppressed, and biochemical evidence of deficiency was present even if some fat was provided by oral nutrition in combination with parenteral feeding (Stein et al. 1980). By the mid-1970s, however, intravenous fat emulsions had become generally available, and it was accepted that a sufficient supply of linoleic acid was one of the important factors to be considered in every case of parenteral nutrition (Wolfram et al. 1978).

The Essentiality of Linolenic Acid

The earliest studies of unsaturated fatty acids (in common and chemical names; see Table IV.C.1.1) showed that both linoleic and linolenic acids had beneficial effects upon the clinical signs of deficiency in rats. Linoleic and arachidonic acids cured the deficiency's symptoms of growth retardation, skin lesions, and excessive water consumption, whereas linolenic acid only cured growth retardation (Burr 1942). Attempts were subsequently made by a group of investigators at Berkeley to produce linolenic acid deficiency by maintaining rats for 3 generations on a diet lacking *n*-3 fatty acids. Levels of *n*-3 in tissues became very low, but small amounts remained and the rats showed no abnormality in growth, reproduction, or appearance (Tinoco et al. 1971). The Berkeley investigators then used radioactive carbon-labeled linolenic acid to trace the impact on dietary fat sources when linolenic acid was converted to docosahexaenoic acid (Poovaiah, Tinoco, and Lyman 1976). Measured in liver phospholipids, the radioactivity was recovered as 20:5*n*-3 and 22:6*n*-3. The dietary fat supplements containing *n*-6 fatty acids (linoleic and arachidonic acids) reduced the conversion of 20:4*n*-3 to 20:5*n*-3 (desaturation), whereas the *n*-3 supplements (18:3*n*-3 and 22:6*n*-3) reduced the conversion of 20:5*n*-3 to 22:5*n*-3 (elongation) (Figure IV.C.1.1). The researchers concluded that 22:6*n*-3 may control its own formation by regulating elongation.

Table IV.C.1.1. *Unsaturated fatty acids*

Chemical description	Chemical name	Common name
18:1n-9	9-octadecenoic	Oleic, 18:1ω9
18:2n-6	9,12-octadecadienoic	Linoleic, 18:2ω6
18:3n-6	6,9,12-octadecatrienoic	Gamma-linolenic, 18:3ω6
18:3n-3	9,12,15-octadecatrienoic	Alpha-linolenic, 18:3ω3
20:3n-6	8,11,14-eicosatrienoic	Dihomogamma-linolenic, 20:3ω6
20:4n-6	5,8,11,14-eicosatetraenoic	Arachidonic, 20:4ω6
20:5n-3	5,8,11,14,17-eicosapentaenoic	Timnodonic EPA, 20:5ω3
22:4n-6	7,10,13,16-docosatetraenoic	Adrenic, 22:4ω6
22:5n-6	4,7,10,13,16-docosapentaenoic	22:5ω6
22:5n-3	7,10,13,16,19-docosahexaenoic	Cervonic DHA, 22:5ω3

Evidence, though inconclusive, that linolenic acid is essential was gathered by examining tissues of rats depleted of the n-3 family. Brain and retinal tissue retained docosahexaenoic acid tenaciously through two generations of rat growth (Tinoco, Miljanich, and Medwadowski 1977; Tinoco et al. 1978), but prolonged deprivation of n-3 fatty acids resulted in reduced visual acuity in infant monkeys and defective electroretinographic responses in monkeys and rats (Neuringer, Anderson, and Connor 1988). As was the case with linoleic acid, patients requiring intravenous feeding have been important in proving that linolenic acid is essential in humans. Patients on TPN, observed by K. S. Bjerve (1989) from 1987 to 1989, experienced scaly and hemorrhagic dermatitis, hemorrhagic folliculitis of the scalp, growth retardation, and impaired

wound healing. But the addition of different oils showed that 1.0 to 1.2 percent of energy as linolenate was necessary to obtain a normal concentration of n-3 fatty acids and relieve the symptoms.

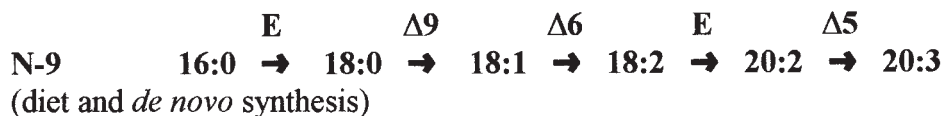
The Discovery of Prostaglandins

Meanwhile, in Sweden, Sune Bergstrom and his colleagues (1962) had determined the structure of a new class of compounds that had been isolated from the vesicular glands of sheep and were named prostaglandins. The subsequent discovery that essential fatty acids were the natural precursors of the prostaglandins was made simultaneously by Bergstrom's team, by a group in Holland, and by another at the Upjohn Company in the United States (Bergstrom 1972). That essential fatty acid deficiency could affect prostaglandin functions was shown by Bergstrom and L.A. Carlson (1965).

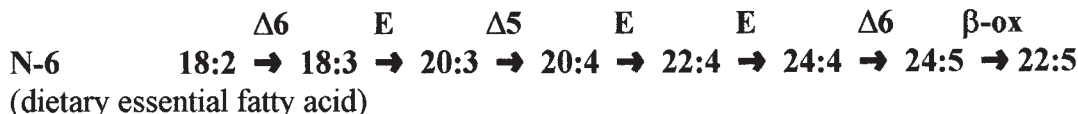
The first international conference on prostaglandins was held in 1972, and Bergstrom, who opened the meeting, commented on the difficulty of evaluating analytical methods for prostaglandins. At that time, bioassay and gas chromatography-mass spectrometry (GC-MS) were being used. Radioimmunoassay was quite new and considered unreliable, as were bioassays. Bergstrom (1973) called for intensive discussion of the analytical questions. Daniel H. Hwang, at the time a doctoral student in nutrition, proposed that in order to study dietary effects on prostaglandin status, a reliable radioimmunoassay should be developed. GC-MS was not sensitive enough and was too time-consuming and expensive to be used for analysis of large numbers of biological samples.

Very little was known then about the metabolism of prostaglandins. The short half-life, explosive synthe-

OLEIC ACID



LINOLEIC ACID



LINOLENIC ACID

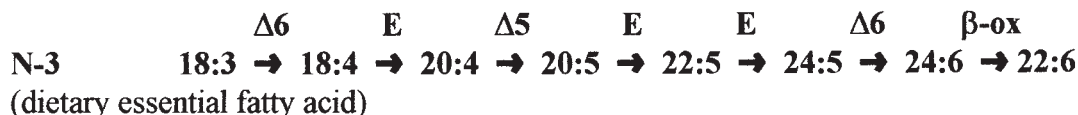


Figure IV.C.1.1. Desaturation (Δ), elongation (E), and chain shortening (β -ox, beta oxidation) of families of unsaturated fatty acids.

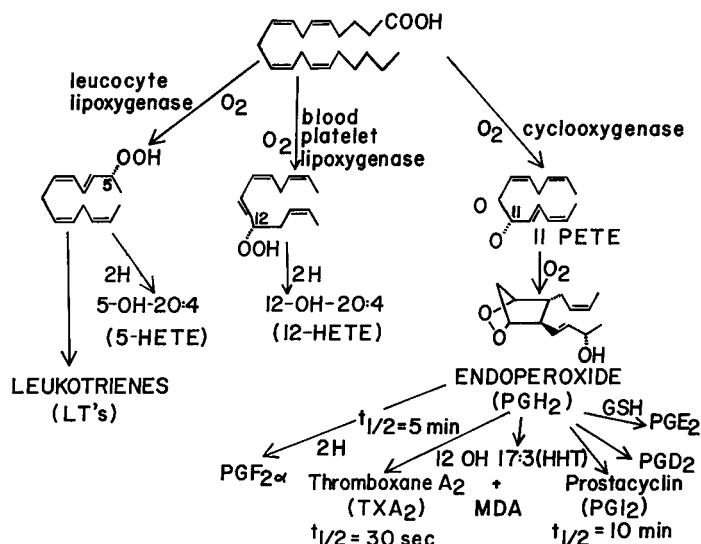


Figure IV.C.1.2. Conversion of arachidonic acid (eicosatetraenoic acid) into eicosanoids. Lipxygenases convert 20:4n-6 to hydroxyperoxyeicosatrienoic acid (HETE) that is converted to leukotrienes. Cyclooxygenases convert 20:4n-6 to peroxycyclo-octatrienoic acid (PETE) that converts to endoperoxide (PGH₂) that is converted to prostaglandins (PG), thromboxane (TX), or prostacyclin (PGI₂). The half-life (*t*_{1/2}) of thromboxane in vivo is 30 seconds and that of prostacyclin is about 10 minutes.

sis in response to trauma, and tissue specificity were yet to be discovered. Serendipitously, Hwang and colleagues (1975) chose to analyze blood serum from rats that had been anesthetized, and these experiments were the first to apply the discovery of prostaglandins to the understanding of essential fatty acid functions and requirements. Rats were fed diets containing corn oil or beef tallow as the source of fat, which showed that there was a positive effect of corn oil (containing linoleic acid) on the synthesis of PGE₁ and PGF_{2α}.

Subsequently, the laboratory of Melvin M. Mathias and Jacqueline Dupont (1985) demonstrated a biphasic response of prostaglandin synthesis to the dietary content of linoleic acid. The response to 0 to 2 percent of energy from linoleate was an increase in prostaglandin synthesis; with 2 to 5 percent of energy there was a decrease, and above 5 percent of energy there was a gradual increase up to the maximum content of linoleic acid fed (27 percent of energy). The 0 to 2 percent energy response to linoleate was associated with a disappearance of eicosatrienoic acid. Responses to a higher arachidonic acid concentration in serum (Fig. IV.C.1.2).

The importance of the relation of dietary linoleate to eicosanoid synthesis meant the addition of a functional measurement to the earlier indications of essential fatty acid deficiency, both clinical and biochemical. This functional indicator of essential fatty acid requirement suggested that 5 to 10 percent of linoleate is desirable (Dupont and Dowd 1990).

Modern History

Excessive extradermal water loss, one of the first symptoms described of essential fatty acid deficiency, has

been explained. A major specific function of linoleate is in skin ceramides, where the linoleate is incorporated into acylglucosylceramides and acylceramides (Hansen and Jensen 1985). Linoleate is the only fatty acid substantially incorporated into these sphingolipids. On another front, the details of control of elongation and desaturation and interrelationships among the families of fatty acids are still being investigated (Cook et al. 1991; Cunnane et al. 1995; Sprecher et al. 1995).

Since the 1970s, an enormous literature about prostaglandin metabolism has accumulated. The term "eicosanoids" was introduced in 1980 to describe the class of substances having 20 carbon atoms derived from n-6 and n-3 fatty acids. Because eicosanoids are regulators of a large array of physiological functions, there are many possible manifestations of deficiency of their precursors. The attention given to the very long-chain fatty acids from marine animals (fish oils) has created another large body of literature about the interplay between metabolism of omega-3 (n-3) and omega-6 (n-6) families of fatty acids. The competition between the two families for enzymes of chain elongation, desaturation, and conversion to active metabolites first demonstrated by Holman (1964) has introduced additional complexity into the attempt to define the dietary requirements for the two families. We know that a source of 18 carbon n-3 and n-6 fatty acids is a dietary necessity, but whether there may be health benefits from consumption of the longer chain products – 20:4n-6, 22:5n-3, and 22:6n-3 – is a current topic of intensive research. Infant requirements for brain development are important considerations for assuring appropriate recommendations for infant formulas and supplements (Carlson et al. 1993).

A broad range of ongoing research is aimed at defining the functions of n-3 fatty acids, and a defi-

ciency has been associated with the function of rhodopsin (Bush et al. 1994). Many aspects of brain and behavioral development are linked to the availability of linolenic acid and its products (Neuringer, Reisbick, and Janowsky 1994). The effects of the ratio of dietary linoleic to linolenic acids is the subject of current research (Jensen et al. 1996), and immune functions and cell signaling are also current topics of study of fatty acid functions and metabolism (Hayek et al. 1997; Hwang in press).

Summary

The evolution of an understanding of essential fatty acids has progressed through several scientific developmental phases. The earliest was the ability to prepare diets of sufficient purity to exclude lipids, which led to the conclusion that some fat was essential. The substantive proof of a human requirement – that is, that the fatty acids were essential to the diet – had to await the advent of total parenteral feeding. In about the same era, eicosanoids were discovered, which opened a new world for understanding the functions of essential fatty acids. With the tools of science today, the profound participation of fatty acids in all aspects of cellular life and function should provide great excitement as well as great challenges to all scientists interested in the expansion of the science of nutrition.

Jacqueline L. Dupont

Bibliography

- Alexander, C. S., and L. Ziene. 1961. Fat infusions. Toxic effects and alterations in fasting serum lipids following prolonged use. *Archives of Internal Medicine* 107: 514.
- Bergstrom, S. 1972. Introduction, International Conference on Prostaglandins, 1972. In *Advances in the Biosciences* 9: 1–5.
- Bergstrom, S., and L. A. Carlson. 1965. Lipid mobilization in essential fatty acid deficient rats. *Acta Physiologica Scandinavica* 64: 479–80.
- Bergstrom, S., F. Dressler, R. Ryhage, et al. 1962. Prostaglandin and related factors. VIII. The isolation of two further prostaglandins from sheep prostate glands. *Arkiv Kemi* 19: 563–7.
- Bivins, B. A., R. P. Rapp, K. Record, et al. 1980. Parenteral safflower oil emulsion (Liposyn 10%). Safety and effectiveness in treating or preventing essential fatty acid deficiency in surgical patients. *Annals of Surgery* 203: 307–15.
- Bjerve, K. S. 1989. N-3 fatty acid deficiency in man. *Journal of Internal Medicine* 225 (Supplement 1): 171–5.
- Burr, G. O. 1942. Significance of the essential fatty acids. *Federation Proceedings* 1: 224–33.
- Burr, G. O., and M. M. Burr. 1929. A new deficiency disease produced by the rigid exclusion of fat from the diet. *The Journal of Biological Chemistry* 82: 345–67.
1930. On the nature and the role of fatty acids essential in nutrition. *The Journal of Biological Chemistry* 86: 587–621.
- Bush, R. A., A. Malnoe, C. E. Reme, and T. P. Williams. 1994. Dietary deficiency of N-3 fatty acids alters rhodopsin content and function in rat retina. *Investigative Ophthalmology and Visual Science* 35: 91–100.
- Carlson, S. E., S. H. Werkman, P. G. Rhodes, and E. A. Tolley. 1993. Visual acuity development in healthy preterm infants: Effect of marine-oil supplementation. *American Journal of Clinical Nutrition* 58: 35–42.
- Cook, H. W., D. M. Byers, F. B. St. C. Palmer, et al. 1991. Alternate pathways in the desaturation and chain elongation of linolenic acid, 18:3(n-3), in cultured glioma cells. *Journal of Lipid Research* 32: 1265–73.
- Cornbleet, T. 1935. Use of maize oil (unsaturated fatty acids) in the treatment of eczema. *Archives of Dermatology and Syphilology* 31: 224–6.
- Cunnane, S. C., M. A. Ryan, K. S. Craig, et al. 1995. Synthesis of linoleate and alpha-linolenate by chain elongation in the rat. *Lipids* 30: 781–3.
- Dudrick, S. J., D. W. Wilmore, H. M. Vars, and J. E. Rhoads. 1968. Long-term total parenteral nutrition with growth, development, and positive nitrogen balance. *Surgery* 64: 134–42.
- Dupont, J., and M. K. Dowd. 1990. Icosanoid synthesis as a functional measurement of essential fatty acid requirement. *Journal of the American College of Nutrition* 9: 272–6.
- Evans, H. M., and G. O. Burr. 1927a. A new dietary deficiency with highly purified diets. *Proceedings of the Society for Experimental Biology and Medicine* 24: 740–3.
- 1927b. A new dietary deficiency with highly purified diets. II. Supplementary requirement of diet of pure casein, sucrose, and salt. *Proceedings of the Society for Experimental Biology and Medicine* 25: 41–8.
1928. A new dietary deficiency with highly purified diets. III. The beneficial effect of fat in diet. *Proceedings of the Society for Experimental Biology and Medicine* 25: 390–7.
- Evans, H. M., and S. Lepkovsky. 1932a. Vital need of the body for certain unsaturated fatty acids. I. Experiments with fat-free diets in which sucrose furnishes the sole source of energy. *The Journal of Biological Chemistry* 96: 143–56.
- 1932b. Vital need of the body for certain unsaturated fatty acids. II. Experiments with high fat diets in which saturated fatty acids furnish the sole source of energy. *The Journal of Biological Chemistry* 96: 157–64.
- 1932c. Vital need of the body for certain unsaturated fatty acids. III. Inability of the rat organism to synthesize the essential unsaturated fatty acids. *The Journal of Biological Chemistry* 99: 231–4.
- Evans, H. M., S. Lepkovsky, and E. A. Murphy. 1934a. Vital need of the body for certain unsaturated fatty acids. IV. Reproduction and lactation upon fat-free diets. *The Journal of Biological Chemistry* 106: 431–40.
- 1934b. Vital need of the body for certain unsaturated fatty acids. V. Reproduction and lactation upon diets containing saturated fatty acids as their sole source of energy. *The Journal of Biological Chemistry* 106: 441–4.
- 1934c. Vital need of the body for certain unsaturated fatty acids. VI. Male sterility on fat-free diets. *The Journal of Biological Chemistry* 106: 445–50.
- Fulco, A. J., and J. F. Mead. 1959. Metabolism of essential fatty acids. VIII. Origin of 5,8,11-eicosatrionic acid in the fat-deficient rat. *The Journal of Biological Chemistry* 234: 1411–16.
- Hallberg, D., O. Schuberth, and A. Wretling. 1966. Experimental and clinical studies with fat emulsion for intravenous nutrition. *Nutra Dieta* 8: 245.

- Hansen, A. E. 1933. Serum lipid changes and therapeutic effect of various oils in infantile eczema. *Proceedings of the Society for Experimental Biology and Medicine* 31: 160-1.
- Hansen, A. E., D. J. D. Adam, A. N. Boelsche, et al. 1957. Manifestations of fat deficiency in infants. *Federation Proceedings* 16: 387.
- Hansen, A. E., and H. F. Wiese. 1951. Fat in the diet in relation to nutrition of the dog: I. Characteristic appearance and gross changes of animals fed diets with and without fat. *Texas Reports on Biology and Medicine* 9: 491-515.
1954. Essential fatty acids and human nutrition: II. Serum level for unsaturated fatty acids in poorly-nourished infants and children. *Journal of Nutrition* 52: 367-74.
- Hansen, H. S., and B. Jensen. 1985. Essential function of linoleic acid esterified in acylglucosylceramide and acylceramide in maintaining the epidermal water permeability barrier. Evidence from feeding studies with oleate, linoleate, arachidonate, columbinatate and α -linolenate. *Biochimica and Biophysica Acta* 834: 357-63.
- Hayek, M. G., C. Mura, D. Wu, et al. 1997. Enhanced expression of inducible cyclooxygenase with age in murine macrophages. *Journal of Immunology* 159: 2445-51.
- Holman, R. T. 1964. Nutritional and metabolic interrelationships between fatty acids. *Federation Proceedings* 23: 1062-7.
1973. Essential fatty acid deficiency in humans. In *Dietary lipids and postnatal development*, ed. C. Galli, 127-43. New York.
- Hwang, D. H., and S. H. Rhee. 1999. Receptor-mediated signaling pathways: Potential targets of modulation by dietary fatty acids. *American Journal of Clinical Nutrition* 70: 545-56.
- Hwang, D. H., M. M. Mathias, J. Dupont, and D. L. Meyer. 1975. Linoleate enrichment of diet and prostaglandin and prostaglandin metabolism in rats. *Journal of Nutrition* 105: 995-1002.
- Jensen, C. L., H. Chen, J. K. Fraley, et al. 1996. Biochemical effects of dietary linoleic/ α -linolenic acid ratio in term infants. *Lipids* 31: 107-13.
- Mathias, M. M., and J. Dupont. 1985. Quantitative relationship between dietary linoleate and prostaglandin (icosanoid) biosynthesis. *Lipids* 20: 791-801.
- Mayer, J., and S. D. Hanson. 1960. Michel Eugene Chevreul - A biographical sketch. *Journal of Nutrition* 72: 3-7.
- McAmis, A. J., W. E. Anderson, and L. B. Mendel. 1929. Growth of rats on "fat-free" diets. *The Journal of Biological Chemistry* 82: 247-62.
- Neuringer, M., G. J. Anderson, and W. E. Connor. 1988. The essentiality of n-3 fatty acids for the development and function of the retina and brain. *Annual Review of Nutrition* 8: 517-41.
- Neuringer, M., S. Reisbick, and J. Janowsky. 1994. The role of n-3 fatty acids in visual and cognitive development: Current evidence and methods of assessment. *Journal of Pediatrics* 125: S39-47.
- Niswender, G. D., A. M. Akbar, and T. M. Nett. 1975. Use of specific antibodies for quantification of steroid hormones. *Methods in Enzymology* 36, Part A: 16-34.
- Paulsrud, J. R., L. Pensler, C. F. Whitten, et al. 1972. Essential fatty acid deficiency in infants by fat-free intravenous feeding. *American Journal of Clinical Nutrition* 25: 897-904.
- Poovaiah, B. P., J. Tinoco, and R. L. Lyman. 1976. Influence of diet on conversion of $^{14}\text{C}_1$ -linolenic acid to docosahexaenoic acid in the rat. *Lipids* 11: 194-202.
- Reiser, R., B. Gibson, M. J. Carr, and B. G. Lamp. 1951. The synthesis and interconversions of polyunsaturated fatty acids by the laying hen. *The Journal of Nutrition* 44: 159-76.
- Rieckehoff, I. G., R. T. Holman, and G. O. Burr. 1949. Polyethenoid fatty acid metabolism. Effect of dietary fat on polyethenoid fatty acids in rat tissues. *Archives of Biochemistry* 20: 331-40.
- Sprecher, H., D. L. Luthria, B. S. Mohammed, and S. P. Baykousheva. 1995. Reevaluation of the pathways for the biosynthesis of polyunsaturated fatty acids. *Journal of Lipid Research* 36: 2471-7.
- Stein, T. P., G. P. Buzby, W. C. Hargrove, III, et al. 1980. Essential fatty acid deficiency in patients receiving simultaneous parenteral and oral nutrition. *Journal of Parenteral and Enteral Nutrition* 4: 343-5.
- Steinberg, G., W. H. Slaton, D. R. Howton, and J. F. Mead. 1956. Metabolism of essential fatty acids. IV. Incorporation of linoleate into arachidonic acid. *The Journal of Biological Chemistry* 220: 257-64.
- Tinoco, J., R. Babcock, R. Hincenbergs, et al. 1978. Linolenic acid deficiency: Changes in fatty acid patterns in female and male rats raised on a linolenic acid-deficient diet for two generations. *Lipids* 13: 6-17.
- Tinoco, J., P. Miljanich, and B. Medwadowski. 1977. Depletion of docosahexaenoic acid in retinal lipids of rats fed a linolenic acid-deficient, linoleic-acid containing diet. *Biochimica and Biophysica Acta* 486: 575-8.
- Tinoco, J., M. A. Williams, I. Hincenbergs, and R. L. Lyman. 1971. Evidence for nonessentiality of linolenic acid in the diet of the rat. *Journal of Nutrition* 101: 937-46.
- Wapnick, S., D. A. Norden, and D. J. Venturas. 1974. Essential fatty acid deficiency in patients with lesions of the gastrointestinal tract. *Gut* 15: 367-70.
- White, H. B., M. D. Turner, A. C. Turner, and R. C. Miller. 1973. Blood lipid alterations in infants receiving intravenous fat-free alimentation. *The Journal of Pediatrics* 83: 305-13.
- Widmer, C., and R. T. Holman. 1950. Polyethenoid fatty acid metabolism. II. Deposition of polyunsaturated fatty acids in fat-deficient rats upon single fatty acid supplementation. *Archives of Biochemistry* 25: 1-12.
- Wiese, H. F., R. H. Gibbs, and A. E. Hansen. 1954. Essential fatty acids and human nutrition: I. Serum level for unsaturated fatty acids in healthy children. *Journal of Nutrition* 52: 355-65.
- Wiese, H. F., and A. E. Hansen. 1953. Semimicromethod for unsaturated fatty acids of blood serum. *Journal of Biological Chemistry* 202: 417.
- Wiese, H. F., A. E. Hansen, and D. J. D. Adams. 1958. Essential fatty acids in infant nutrition. I. Linoleic acid requirement in terms of serum di-, tri- and tetraenoic acid levels. *Journal of Nutrition* 66: 345-60.
- Wolfram, G., J. Eckart, B. Walther, N. Zollner. 1978. Factors influencing essential fatty acid requirement in total parenteral nutrition (TPN). *The Journal of Parenteral and Enteral Nutrition* 2: 634-9.

IV.C.2 🐾 Proteins

The word "protein" was coined by Jöns Jakob Berzelius in 1838. For the previous 150 years, however, there had been the concept of an "animal substance," slight variants of which were thought to make up muscles, skin, and blood. In each form the substance was initially believed to be gluey. But it

turned into hard, hornlike material when heated and became foul-smelling when kept under moist, warm conditions, giving off an alkaline vapor. This contrasted with the properties of starch and sugar and most whole plants that went to acid during damp, warm storage.

For people interested in nutrition, the obvious question was: "How does the animal kingdom, which as a whole lives on the plant kingdom, convert what it eats into the apparently very different animal substance?" Humans were, of course, included in the animal kingdom and assumed to have essentially the same nutritional system as animals. Some eighteenth-century discoveries threw light on the problem.

In 1728, the Italian scholar Jacopo Beccari announced that he had discovered the presence of a material with all the characteristics of "animal substance" in white wheat flour. When he wetted the flour to make a ball of dough, then washed and kneaded it in water, the fine, white starchy particles washed out. What remained was a sticky pellet of gluten, which, if its origin were unknown, would be judged animal in nature. Beccari concluded that the presence of this portion of preformed "animal substance" made wheat particularly nutritive. Wheat flour, as a whole, did not show animal properties because the greater quantity of starch overwhelmed the reactions of the gluten.

Nitrogen in Nutrition

Later in the eighteenth century, with the development of the new chemistry, the main elements were identified, and ammonia, the "volatile alkali," was shown to be a compound of nitrogen and hydrogen. Gluten was also found to contain nitrogen, in common with animal tissues, whereas starches, fat, and sugars did not.

At first it was thought that the process of animal digestion and nutrition must consist of the combining of nutrients in plant foods with atmospheric nitrogen in order to "animalize" them. In particular, it seemed that this theory might explain the slow digestion process and large storage stomachs in ruminant animals. However, further work in France made this appear less likely.

First, François Magendie reported in 1816 that dogs failed to survive for more than a few weeks on foods like fats and sugars that contained no nitrogen. Then, in the 1830s, Jean Boussingault showed that the nitrogen present in the hay and potatoes eaten by a cow was enough to balance the quantities present in the milk it secreted together with its regular daily nitrogen losses. There was, therefore, no need to suppose that atmospheric nitrogen was involved in animal nutrition. But because of the importance of nitrogen in nutrition, Boussingault concluded that plant foods should be valued in terms of their relative nitrogen contents. Thus, he believed that dry beans, with

roughly twice the nitrogen content of grains, had twice their nutritional value.

By this time, further work on the composition of plants had shown that although they all contained nitrogenous compounds, most of them, unlike wheat gluten, were soluble in water, yet could be precipitated by heat or acid. In 1838, Gerritt Mulder, a Dutch physician who had taught himself chemical analysis, published a claim that all the important "animal substances" he had analyzed had the same basic formula, corresponding to 40 atoms of carbon, 62 of hydrogen, 10 of nitrogen and 12 of oxygen, which can be expressed more simply as $C_{40}H_{62}N_{10}O_{12}$. They differed in their properties only because they had different numbers of atoms of sulfur and/or phosphorus adhering to them. He sent his paper to the Swedish chemical authority, Jacob Berzelius, who replied that this was a most important discovery of the "fundamental or primary substance of animal nutrition" and that this substance deserved to be called "protein" after the Greek god Proteus.

The leading German organic chemist, Justus Liebig, confirmed Mulder's finding and went on to argue that, from a chemical point of view, it was the plant kingdom alone that had the power of making protein. Animal digestion only loosened the association between their molecules to make them soluble and absorbable into the bloodstream and immediately ready for deposit into the animal system. The leading French scientists accepted this view but added that vegetable oils and carbohydrates were also required. Their combustion was needed within the animal body to maintain animal heat.

Protein as a Muscle Fuel

Liebig, although he had himself done no physiological work, developed a whole series of dogmatic statements as to the functions of nutrients in the body. He believed that the energy needed for the contraction of muscles came solely from the breakdown of some of their own protein, which was then immediately decomposed further, with the nitrogenous portion appearing as urea in the urine. A subject's requirement for protein was, therefore, proportional to his or her performance of physical work. The role of fats and sugars was merely to protect living tissues (which reacted with oxygen that penetrated them) from the danger of oxygen damage. Consequently, protein was the only true nutrient.

Liebig's views were accorded great weight, although there were many grounds on which they could be criticized. For example, in 1862, Edward Smith, a physician and physiologist who had been studying the health and diet of the inmates of London prisons, reported a study of factors influencing the daily output of urea. Prisoners who ate the same rations each day and engaged in hard labor three days per week were found to excrete almost the same quantity of urea on the day (and following night) of

the labor as on the days when not laboring. However, the labor caused greatly elevated carbon dioxide output in the breath. The main factor influencing urea production appeared to be the amount of protein eaten in the previous 24 hours.

In 1865, Adolf Fick and Johannes Wislicenus, on the faculty of a Swiss university, followed up these findings. They put themselves on a protein-free diet for 24 hours and ascended almost 2,000 meters on a path to the summit of a convenient mountain. They calculated the amount of work done during the ascent and measured the amount of nitrogen in the urine they excreted. From this they calculated that they had each metabolized approximately 37 grams (g) of protein. Their friend in England, Edward Frankland, now calculated that the metabolism of protein yielded 4.37 kilocalories per gram.

By this time the principle of the “conservation of energy” had been accepted, and James Joules had estimated that 1 kilocalorie was equivalent to 423 kilogram-meters (kg-m) of mechanical work against the force of gravity. The energy released from the protein was, therefore, equivalent to some 68,000 kg-m. However, the net work required to lift each scientist up the mountain was approximately 140,000 kg-m, about twice as much. And further work has shown that muscles operate at something like 25 percent efficiency, so that four times the minimal theoretical amount of fuel is required. The conclusion was, therefore, that the energy required for muscular effort does not come primarily from protein but from dietary fats and carbohydrates.

Although Liebig's grand scheme had been discredited, German workers, in particular, continued to maintain that a high-protein intake was desirable to maintain both physical and nervous energy. They argued this on the grounds that people from countries where the diet was largely vegetarian and low in protein lacked “get-up-and-go,” and that wherever people were unrestrained by poverty and could eat what they wished, they chose a high-meat, high-protein diet. The first U.S. government standards, issued at the end of the 1800s by Wilbur Atwater, the Department of Agriculture's nutrition specialist, followed the same line in recommending that physically active men should eat 125 g of protein per day.

Such a notion did not go unchallenged, however. From 1840 on, there had been a vegetarian “school” in the United States, which argued that eating meat was overstimulating and conducive first to debauchery and then to exhaustion of the irritated tissues. John Harvey Kellogg (cofounder of the family's breakfast-food enterprise) believed that meat and other sources of excessive protein in the diet could putrefy in the large intestine, resulting in autointoxication. These ideas were regarded by the scientific establishment as unscientific and not meriting attention. However, in 1902 a serious challenge to the “high-protein” school was mounted by Russell Chittenden, professor of physiological chemistry at Yale.

Chittenden had six of his colleagues, a dozen army corpsmen, and a group of Yale's athletes spend approximately six months on diets containing no more than one-half of the Atwater standard for protein. These men all remained healthy and vigorous in mind and body. Chittenden concluded, in his account published in 1904, that such diets were not only adequate but preferable because they put less strain on the kidney to cope with the excretion of both urea and the less soluble uric acid.

His findings stimulated an active debate among medical men. Most believed that Chittenden's findings were still too limited to recommend wholesale dietary changes. For example, the subjects in his study had not been subjected to sudden stresses or to periods of inadequate feeding in which they had to rely on reserves. Moreover, experiments with dogs kept on low-protein diets had revealed that although they remained healthy and in nitrogen balance for a time, they eventually weakened and died. Chittenden, however, followed up this line of work with dogs in his own laboratory and concluded that it was not lack of protein that was responsible for long-term problems with some diets but a lack of one or more unknown trace nutrients. This conclusion constituted one of the stimuli for the work that dominated nutritional studies for the next 40 years and revealed the existence of the vitamins.

Amino Acids in Nutrition

By 1905, another question gaining prominence was whether the proteins that had now been isolated from many foods could all be considered equivalent in nutritional value. It had long been known that gelatin, obtained by autoclaving bones, would not support growth in dogs, even though it had the same nitrogen content as ordinary tissue proteins. But because it displayed some physical differences, such as remaining soluble in boiling water, it had been set aside from the “protein” classification.

Advances in the study of proteins required a better knowledge of their composition. That they did not diffuse through fine-pored membranes showed them to be large molecules. But during digestion, their physical properties changed dramatically. With the isolation of the digestive agent “pepsin” from the stomach walls of slaughtered animals, and then of “trypsin” from pancreatic juice, the process could be studied in more detail.

As early as the 1860s, workers had been surprised to discover the presence of “leucine” and “tyrosine” in digests of protein with pancreatic juice. These two compounds were already well known. They were recoverable from the product of boiling proteins with sulfuric acid and had been shown to be “amino acids” – meaning that they contained both an acid group and a basic one and were relatively small molecules, each with less than 25 atoms. However, the procedure yielding them seemed so severe as to have no relation

to the mild conditions of the digestive system, and it was thought that the compounds might well have been produced by the hot and strong acid conditions.

Gradually, by *in vitro* digestion, and milder acid or alkali refluxing of proteins, a whole range of amino acids were recovered, and crude methods were developed to analyze the quantities of each that were present. Also, evidence accumulated that the compounds actually absorbed through the gut wall after digestion were simple amino acids. Early investigators had felt it unlikely that nature would employ such a system because it seemed extremely wasteful to break proteins down, only to rebuild the same compounds within the animal. Or were they the same compounds?

Some of the first experiments comparing the nutritional values of different proteins were carried out by Lafayette Mendel (from Chittenden's group) and the plant chemist Thomas Osborne. They found that young rats would grow if given a diet of fat, carbohydrates, minerals, crude vitamin concentrates, and purified casein (a milk protein). However, with zein (a protein from corn) as the protein source, the rats did not grow unless the diet was fortified with *both* lysine and tryptophan. Chemical analysis had already indicated that zein lacked these two amino acids, and thus these two were characterized as "essential amino acids," meaning that they were essential or indispensable in the diet of growing rats. The results also indicated that animals used amino acids to build their own body proteins.

After 20 years of further experiments of this kind, W. C. Rose and his colleagues were able to obtain good growth in rats with diets containing *no* protein, but just a mixture of amino acids in its place. Table IV.C.2.1 summarizes their findings about the 20 amino acids present in animal proteins, some of which were indispensable ("essential") and some of which the rat could make for itself ("nonessential") if they were not supplied in its diet. Further work led to the development of values for the quantity of each indispensable amino acid that rats required for optimal growth.

Another group of studies compared the relative values of different protein sources (or of mixtures) for the support of growth in rats. The mixed proteins from individual vegetable foods (grains, beans, and so forth) all supported some growth, but not to quite the same extent as the mixed proteins in milk, meat, or eggs. The first limiting amino acid (meaning that this was the amino acid that increased growth when added as a single supplement) in most grains was lysine. This was to be expected in view of the growing rat's known requirement for lysine and the low analytical value of lysine in grains. The corresponding first limiting amino acid in most beans and peas was found to be methionine. Because the two classes of materials had different deficiencies, one would expect a mixture of grains and legumes to support better growth in rats, and this has been confirmed.

Table IV.C.2.1. *Reproduction of the final summary of the rat's requirements for amino acids, as determined by Rose and his colleagues in 1948*

Classification of amino acids with respect to their growth effects in the rat	
Essential	Nonessential
Lysine	Glycine
Tryptophan	Alanine
Histidine	Serine
Phenylalanine	Cystine*
Leucine	Tyrosine†
Isoleucine	Aspartic acid
Threonine	Glutamic acid‡
Methionine	Proline‡
Valine	Hydroxyproline
Arginine§	Citrulline

*Cystine can replace about one-sixth of the methionine requirement, but has no growth effect in the *absence* of methionine.

†Tyrosine can replace about one-half of the phenylalanine requirement, but has no growth effect in the *absence* of phenylalanine.

‡Glutamic acid and proline can serve individually as rather ineffective substitutes for arginine in the diet. This property is not shared by hydroxyproline.

§Arginine can be synthesized by the rat, but not at a sufficiently rapid rate to meet the demands of *maximum* growth. Its classification, therefore, as essential or nonessential is purely a matter of definition.

Source: Rose, Oesterling, and Womack (1948). Reprinted in Carpenter 1994, p. 133.

Human Requirements

Rats, however, although useful as models, differ from humans (even in this context) in important ways. Humans spend most of their lives as adults, not growing at all but needing protein just for "maintenance." And in childhood, human growth is extremely slow compared to the growth of rats. Thus, we take something like six months to double our birth weight, which a young rat does in a few days. And at six months, the rat is fully matured, yet the child is still only one-tenth of its mature size. Moreover, although the tissue proteins of rats and humans are similar, hair protein is very different, and the rat has to synthesize proportionally more.

It was necessary, therefore, to discover whether humans needed to be supplied with the same essential amino acids as those needed by rats. But because it was neither practical (nor ethical) to keep young children on what might be inadequate experimental diets for long periods in order to compare their growth rates, the normal method of experimentation was to feed adult volunteers for periods of two weeks or so on diets in which there were mixtures of amino acids in place of protein.

If an essential amino acid were missing, the subject would, within a very few days, show a negative nitrogen balance, meaning that the amount of combined nitrogen found in urine and feces, plus the

smaller estimated quantity rubbed off in skin and hair losses, had exceeded the daily nitrogen intake. Fortunately, no harm seems to come to humans in negative balance for a short period, and bodily reserves refill rapidly on resumption of a complete diet.

The first major finding from this work was that essential and nonessential amino acid needs are the same for humans as for the young rat. However, researchers were surprised to discover how low the quantitative need for each essential amino acid appeared to be in order to maintain nitrogen balance. In fact, the combined total of essential amino acids came to only 16 percent of the total protein requirement, even though they make up about 45 percent of our body proteins. Thus, it seemed that almost any mixture of foods that provided at least the minimum amount of total protein needed would automatically meet adult needs for each essential amino acid. For young children, however, it was felt safer to set a higher standard for amino acids, corresponding more or less to the composition of the proteins in breast milk. For older children, a compromise was adopted in official recommendations, with a pattern midway between that found to be needed for nitrogen balance in adults and that in human milk. These standards are summarized in Table IV.C.2.2.

There have been recent criticisms of the practice of basing standards solely on short-term nitrogen balance experiments, with V. R. Young (1986) and colleagues (1988, 1989) at the Massachusetts Institute of Technology (M.I.T.) arguing that the method itself has sources of error. These researchers have carried out sophisticated studies using diets based on amino acids, with a single essential amino acid labeled with an isotope so that its metabolism can be followed. They concluded that the levels at which the essential amino acids are required, in relation to total protein needs, are quite similar to the levels in which they occur in the body. Even after subjects have had time to adjust to lower intakes, the rate of renewal of body tissues is reduced, which may have adverse effects in a time of stress.

Table IV.C.2.2. *The World Health Organization (1985) estimates of human requirements for protein and selected amino acids, by age*

	Protein (g/kg body weight)	Lysine (mg/g protein)	Methionine + cystine (mg/g protein)
Infant			
(3–4 months)	1.47	103	58
Child (2 years)	1.15	64	27
School child			
(10–12 years)	1.00	44–60	27
Adult	0.88	12	22

As Atwater suggested over a century ago, it is possible that intakes higher than those needed for nitrogen balance could confer some more subtle long-term benefits. However, there are as yet no studies of peoples living for long periods on diets borderline in protein but well served with all other nutrients that would clarify the situation. And in any event, it seems clear that even the higher amino acid levels proposed by the M.I.T. group are being fulfilled by the diets of most people, at least in the developed countries.

The Protein Contribution of Different Foods

The obvious way to express the level of protein in a food is as a percentage of the weight, like “g per 100g.” But such a measurement can be deceptive. For example, it would show ordinary white bread to have nearly 3 times the protein content of cow’s milk because milk is 90 percent water, whereas bread is only about 34 percent water. Alternatively, one could compare the amounts of protein in equal weights of dry matter, but the common nutritional value of the great majority of the dry matter is its contribution of usable energy, whether from carbohydrate, fat, or protein. Thus, nutritionists have found it useful to compare the protein concentration of different foods in relation to their total calorie values. This could be expressed as “g per 100 kcalories,” but it is easier (as protein itself has an average energy value of 4kcal/g) to express the concentration as “protein calories as a percent of total calories” (PCals%).

Although most of the time people have an instinct to eat enough food to meet their energy needs, there is a question of whether this quantity will also include enough protein.

Returning to the comparison of bread and milk, we can make the following comparisons:

	Protein (g)	Energy (kcal)	PCals%
1 slice white bread (32 g)	3	96	12.5
1 cup whole milk (244 g)	8	150	21.3
1 cup skim milk (245 g)	8	86	37.2

In terms of PCals%, milk is richer in protein than bread, meaning that to get the same quantity of protein from bread as from a cup of milk, one would have to consume more total calories. Similarly, it can also be seen that although a cup of whole milk and one of skim milk (with the cream removed) have the same protein content, the PCals% values are very different, with the value for the skim milk being higher. There are equally large differences between different meat preparations, as can be seen in the comparison of a pork chop and a chicken breast:

Table IV.C.2.3. *The typical protein concentrations of a variety of foods (edible portions only) expressed as "protein calories as a percentage of total calories" (PCals%)*

Animal foods		Grain products	
Poached cod	86	Oatmeal	14.5
Roast chicken breast		Whole wheat bread	13.5
without skin	72	White bread	12.5
Stewed rabbit	68	Sweet corn	12
Broiled salmon	50	Brown rice	11
Skim milk	37	Cornmeal	10
Boiled egg	32	White rice	9
Cheddar cheese	25	Sorghum flour	8.5
Fried pork chop	25	Corn flakes, packaged	7.5
Salami (beef and pork)	22		
Whole milk	21	Roots, fruits, etc.	
Liver sausage	17	Baked potato	10
		French fries	5
		Banana	5
Legumes		Sweet potato	4
Tofu (from soy)	43	Plantain	3
Green peas	27	Cassava flour	1
Black beans	26	Fats and sugars	0
Kidney beans	25		
Chickpeas	21		

	Protein (g)	Energy (kcal)	PCals%
Pan-fried pork chop (89 g)	21	334	25
Roasted chicken breast without skin (86 g)	27	142	72

What this shows is that in a fried pork chop, for every 1 g protein (that is, 4 kcal) there are, in addition, 12 kcal from fat, whereas in the roasted chicken breast, 1 g protein is accompanied by only 1.6 kcal from fat. The PCals percentage values for a range of foods are set out in Table IV.C.2.3. These are "average" or "typical" values. Some animal carcasses are fatter than others, and the composition of plant foods can change significantly according to the environment in which the plants are grown, as well as the stage of harvesting. Wheats also differ significantly, with some strains being selected for high or low protein content according to the use for which the flour is marketed.

It is true that animal-product foods are generally higher in protein than plant products and also that people in the more affluent "Westernized" countries eat higher levels of animal products. However, the total protein intake in affluent cultures is not that much larger. The offsetting factor in these cultures is the higher consumption of sugars, fats, and alcoholic beverages, all of which contribute calories but no protein. Moreover, in many developing countries, some kind of beans forms a regular part of the day's food, and they are a rich source of protein. Thus, calculations commonly indicate that diets in both rich and poor countries have mostly between 10.5 and 12.5 percent of their total calories in the form of protein.

The Food and Agriculture Organization of the United Nations (FAO) publishes estimates of the daily food supplies per head that are available in different countries. Here are three examples:

	Total kcal	Protein (g)		% of protein from animals	PCals%	Fat (g)	Sugars (g)
		Veg.	Animal				
U.S.A.	3640	37	72	66	12.0	164	579
Romania	3330	58	44	43	12.3	95	295
Ghana	2200	33	13	28	8.4	43	64

In this comparison, based on recent data, we see that in Romania, a relatively poor European country, the average individual took in only about 60 percent as much animal protein as a counterpart in the United States, but the total protein supply was almost identical. This was because the Romanians ate much less fat and sugar and received correspondingly more calories from grains, which generally have 10 to 13 PCals%. This offsetting, however, breaks down when the staple energy food is not a grain but a starchy root with only 1–3 PCals%, as in West Africa, where cassava is a common staple. The data for Ghana illustrate this. Despite the low fat and sugar intakes, there is still only an overall 8.4 PCals% in the food supply estimated to be available for the average person. Of course, the first "red light" that we see upon looking at the data is the low total calorie intake, which is only 60 percent of the corresponding U.S. value. Not all the U.S. foods are actually consumed of course: There is a great deal of waste, with fat trimmed off meat and stale food thrown out. Conversely, there may be some unrecorded food sources in Ghana. But it is a general finding that, even if there is a good supply of starchy roots, their sheer bulkiness makes it difficult to consume enough to meet energy requirements, particularly for young children, so that neither energy nor protein needs are fully met.

Not surprisingly, West Africa is also the part of the world where the disease kwashiorkor was first studied. It strikes children 1 to 3 years old who appear bloated, though their muscles actually are shrunken. They often have ulcerated and peeling skin and are utterly miserable. Unless treated, they are likely to die. The condition is now thought to be due to a combination of undernutrition (in both protein and energy) with the stress of infections. Recovery can be rapid if the victims are given concentrated food, by stomach tube at first, if necessary. The food mix does not need to be high in protein; mixes with as little as 5 PCals% have proven successful.

Except for children subsisting on bulky and very low protein staples, there seems to be no problem of protein deficiency for people in any culture who can afford to satisfy their calorie needs, unless they are consuming extremely atypical diets. The Recom-

mended Dietary Allowances (RDAs) for protein in the United States are summarized here for three groups, together with the estimated energy needs of individuals in those groups if moderately active:

Population group	Assumed bodyweight (kg)	Protein RDA (g)	Energy needs (kcal)	PCals% required
Children, ages 1-3	13	16	1,300	4.9
Women, ages 25-50	63	50	2,200	9.1
Men, ages 25-50	79	63	2,900	8.7

It is interesting to note that when one calculates the proportions of protein required (PCals%) for each class, the results are unexpected. Traditionally, wives have thought that men, as the “breadwinners” of the family, needed most of the meat, and children at least some extra dairy protein, but as can be seen, it is actually women who are estimated to need the highest proportion of protein in their diet. And for people involved in greater levels of physical activity, all the evidence indicates that their calorie needs increase greatly but not their protein needs, so that the resulting PCals% of their needs is decreased. Put another way, the extra food they need to meet their needs can be of quite low protein content. Similarly, although the protein needs per kg body weight of a 1- to 3-year-old child are 50 percent greater than for an adult, its energy requirement per kg is nearly 200 percent higher so that, once again, the PCals% of its needs are lower.

Returning to the estimated average food supplies in Ghana, it can be seen that the mix is just about at the lower limit for protein. However, the RDA for protein, as for all other nutrients except energy, does include a margin of safety, and the level of physical activity is always higher in developing countries where there is less mechanical transport.

For countries like the United States or Romania, the protein supplies are clearly well above the standard requirement levels. Thus, it follows that a high intake of meat cannot be justified because of the protein that it contributes. In fact, a major concern of late has been that the protein intake of affluent individuals may be undesirably high. Some have problems because their kidneys are inefficient in excreting the urea resulting from protein metabolism. And, even in healthy people, high-protein diets cause increasing urinary losses of calcium. Certainly, this effect is undesirable in a society whose growing percentages of older people contain more and more individuals whose bones have been weakened because of the loss of a considerable proportion of their mineral substance (mostly calcium phosphate). It is now recommended that we not consume more than twice our RDA for protein. This would mean an upper level of 100 g protein for a woman weighing 63 kg (139 lb.).

Kenneth J. Carpenter

Bibliography

- Carpenter, K. J. 1994. *Protein and energy: A study of changing ideas in nutrition*. New York.
- National Research Council. 1989. *Recommended Dietary Allowances*, Tenth edition. Washington, D.C.
- Porter, J. G. W. and B. A. Rolls, eds. 1973. *Proteins in human nutrition*. London.
- Rose, W. C., M. J. Oesterling, and M. Womack. 1948. Comparative growth on diets containing ten and nineteen amino acids, with further observations upon the role of glutonic and aspartic acid. *Journal of Biological Chemistry* 176: 753-62.
- Young, V. R. 1986. Nutritional balance studies: Indicators of human requirement or of adaptive mechanisms. *Journal of Nutrition* 116: 700-3.
- Young, V. R., D. M. Bier, and P. L. Pellet. 1989. A theoretical basis for increasing current estimates of the amino acid requirements in adult man, with experimental support. *American Journal of Clinical Nutrition* 50: 80-92.
- Young, V. R., and P. L. Pellet. 1988. How to evaluate dietary protein. In *Milk proteins: Nutritional, clinical, functional, and technological aspects*, ed. C. A. Barth and E. Schlimmer, 7-36. New York.
- World Health Organization. 1985 *Energy and protein requirements*. Technical Report Series no. 724. Geneva.

IV.C.3 Energy and Protein Metabolism

In conventional scientific usage, when the word metabolism is joined with energy, it takes on a somewhat different meaning than when it is joined with protein. The latter – protein metabolism – usually includes consideration of the biochemical pathways of amino acids, the building blocks of protein, whereas energy metabolism is frequently assumed to include only the specific role of energy without consideration, in any detailed way, of the pathways involved in the breakdown and synthesis of the various carbohydrates and lipids that supply food energy.

In this chapter, the major thrust concerned with energy is an emphasis on historical considerations of understanding and meeting human food energy needs. In the case of protein, the role of amino acids in generating energy, protein quality (including digestibility), and human protein requirements will be given emphasis. Finally, there is some discussion of protein-energy relationships, the problems connected with an excess and a deficit of food energy in the diet, and protein-energy malnutrition.

Food Energy

The most pressing problem for humans throughout their history has been the basic one of securing food to satisfy hunger and food-energy needs. But the fact that the human population in different parts of the world (despite subsisting on different diets), seemed

before about 1900 to experience approximately the same level of health had led some physiologists to believe that all foods were rather similar (McCollum 1957). In fact, this view was rooted in the ancient “many foods–single aliment” concept of Hippocrates and Galen. By the turn of the twentieth century, however, views were rapidly changing, and although knowledge of the specialized roles of amino acids, vitamins, and minerals was hazy and often contradictory, proteins, fats, and carbohydrates (starches and sugars) could be distinguished and analyzed in foods and diets. Thus, with the advances in food analysis and refinements of nutritional needs the “many foods–single aliment” concept was no longer tenable.

Heat and Life

Because the human body became cold after death, it was not surprising that some mechanism of heat production came to be considered synonymous with life at an early stage in human history. Hippocrates (born 460 B.C.) noted in one of his aphorisms:

Growing bodies have the most innate heat; they therefore require the most food for otherwise their bodies are wasted. In old people the heat is feeble and they require little fuel, as it were, to the flame, for it would be extinguished by much. (Lusk 1933: 8)

The Greeks explained the mechanisms of health and illness in terms of the four “humors” in the body. When the body was in health, these humors were in balance, whereas excess or deficiency caused illness. Special foods or drugs could restore the proper balance and, hence, health. It was also observed from the earliest times that when an adult partook of a great deal of food, he or she did not necessarily gain in weight. Hippocrates believed this was due to a constant loss of insensible perspiration and the elimination of heat, which he conceived to be a fine form of matter.

Half a millennium later, Galen (130–200 A.D.) was more specific than Hippocrates in drawing a direct analogy between flame and innate heat, writing that “[t]he blood is like the oil [of a lamp], the heart is like the wick and the breathing lungs an instrument which conveys external motion” (Lusk 1933: 12).

The Middle Ages saw few developments in the area of nutrition, but during the Renaissance, Leonardo da Vinci (1452–1519) again took up the analogy of a flame to refer to life by stating that “where there is life there is heat” and “where flame cannot live no animal can sustain its existence” (Lusk 1933: 18). This remark was later echoed by Robert Boyle in 1660, who wrote that “neither a flame nor an animal can live in a vacuum” (Lusk 1933: 25). A contemporary of Boyle, John Mayow (born 1643), who died at the age of 36, appeared to have an understanding of respiration and the role of blood (that was only rediscovered a century later) when he wrote:

Respiration consists furthermore in the separation of the air by the lungs, and the intermixture with the blood mass of certain particles absolutely necessary to animal life, and the loss by the inspired air of some of its elasticity. The particles of the air absorbed during respiration are designed to convert the black or venous blood into the red or arterial. (Lusk 1933: 35)

Development of Scientific Concepts

Fuel was long understood as burning to produce heat or power, but an understanding of the importance of air in supporting combustion and respiration had to await identification of the gases. Later, when burning was recognized as the combining of fuel with oxygen and the giving off of carbon dioxide and heat, the respiration of animals and humans was accepted as a special case of combustion. Although many still believed that a “vital” or God-given force controlled all life processes, quantitative relationships began to be developed between individual foods and their use in the body to produce predictable amounts of heat.

The concept of heat remained central to the explanatory framework of traditional physiology. Every part of the body had its own innate heat, and the heat in the extremities was continuously being dissipated into the surrounding environment. This heat had to be replenished from some source in the body, which was identified as the heart, and the heat, together with the beating of the heart, generated the “vital spirit” out of air and blood. The vital spirit was, in turn, carried by activated arterial blood.

This view was replaced following the discovery of circulation by William Harvey in 1628. In the process of searching for the origin of body heat, later investigators tried to decide whether the heartbeat was involved with producing it or whether the motion of the blood created friction, and therefore heat, as it passed through the vessels of the body. The lungs were considered by Harvey to represent an important heat-dissipating mechanism that prevented the cardiac generation of heat from exceeding the body’s tolerance.

The Discovery of the Gases

Physiology prospered in the early years of the eighteenth century but during its latter part lost prominence to chemistry. By then, chemists had developed analytical methods to at least begin to determine the ultimate constituents of body fluids and tissues (Kinney 1988).

With the work of Joseph Priestley (1733–1804) oxygen was discovered (1774).

Antoine Laurent Lavoisier, in turn, recognized that carbon dioxide was a compound of carbon and oxygen and further demonstrated that the respiration of animals involved CO_2 . A table from the 1790 English translation of his text *Traité élémentaire de chimie*

indicates that the names oxygen, hydrogen and azote (nitrogen) were replacing several older terms for the gases. By use of an ice calorimeter (a device whereby heat produced from a living animal could be estimated from the amount of ice that was melted), Lavoisier was able to demonstrate that respiration was very slow combustion and that predicted heat agreed well with that which was measured.

Nutritional Experimentation

In the overlapping areas of protein and energy metabolism, French physiologist Jean-Baptiste Boussingault (1802–87), in 1844, performed a series of experiments using a cow fed only foods relatively low in protein (potatoes and beets) to prove that the manufacture of tissue protein did not require the incorporation of nitrogen from the air. A few years later, Carl von Voit and T. L. Bischoff (1860) in Germany confirmed that all the nitrogen in the body could come from the foods eaten and that the condition of nitrogen equilibrium or nitrogen balance could be established when the level of nitrogen intake was held constant. A major controversy developed during this period over the production of animal fat. Justus von Liebig maintained that the formation of fat from sugar was possible in the animal body, whereas other chemists, following the lead of Jean-Baptiste Dumas, maintained that fat in food was the only source of body fat. Boussingault, however, in 1845 proved conclusively that animals could produce fat from carbohydrates (McCollum 1957).

A series of studies that significantly linked protein and energy metabolism were those undertaken by Edward Smith in England in 1857. They demonstrated the effects of different conditions of life on the magnitude of energy requirements of human subjects. His major observation was that there was comparatively little change in the amount of urea (the major protein metabolite) voided during wide variations in the amount of labor performed, whereas the carbon dioxide produced increased in proportion to the amount of exercise. He devised a treadmill experiment with human subjects that provided strong evidence that the energy for muscular work could not have been derived from muscle degradation and almost conclusive evidence that the muscles were deriving energy from nonnitrogenous food. Unfortunately, the high authority of Liebig, who thought energy was derived from muscle degradation, prevented Smith's studies from securing the attention they deserved.

In 1866, however, Adolf Fick and Johannes Wislicenus reported their famous mountain-climbing experiment of the previous year, using themselves as subjects, to confirm Smith's tread-mill studies and also to put Liebig's theory of muscular work to a crucial test. They abstained from protein food for some 17 hours before beginning the ascent of the Faulhorn, a peak in Switzerland with an altitude of 1,656 meters

(5,433 feet). The climb to the summit required 6 hours, and during this time and for another 7 hours afterward, they consumed only carbohydrates and fats. Throughout they collected their urine, which was later analyzed for nitrogen content. During the 13-hour period, Fick produced 5.74 grams of urinary nitrogen and Wislicenus 5.54 grams. They reported:

We can assert from our own experience in the mountain ascent, that in spite of the amount of work and abstinence for 31 hours from albuminous food, neither of us felt the least exhausted. This could hardly have been the case if our muscular force had not been sustained by the non-nitrogenous food in which we partook. (McCollum 1957: 126)

In the same year, E. Frankland (1866) pointed out that these results could be interpreted only by taking into account the amount of energy evolved in the combustion of a unit of muscle substance and by considering the work equivalent of this energy. In addition, he showed that the energy that the body derived from protein was equal to the difference between the total heat of combustion less the heat of combustion of urea, the major substance whereby the nitrogen in protein foods was eliminated. Frankland thus confirmed the conclusions already reached by Smith and by Fick and Wislicenus that muscles work at the expense of energy derived from the oxidation of the nonnitrogenous fats and carbohydrates (McCollum 1957).

The development of the concepts of intermediary metabolism of foodstuffs – that is, the step-by-step conversion of materials to the ultimate products, water and carbon dioxide (CO_2), with the liberation of energy – was a slow intellectual process and had to await the accretion of knowledge of organic chemistry and, above all, the realization of the catalytic nature of the enzymes and their infinite variety (Levine 1978).

Respiration Chambers and Calorimetry

Voit and M. Pettenkofer were leaders in the new studies of respiration and energy balance, at first using small chambers for animals. A much larger apparatus was then constructed in which both the total carbon excretion (breath, urine, and feces) could be measured and the quantity of individual foods actually burned in the human body could be determined. This was a remarkable technological achievement and a significant scientific advance. On the death of Pettenkofer in 1901, Voit (quoted in Kinney 1988) wrote:

Imagine our sensations as the picture of the remarkable process of the metabolism unrolled before our eyes and the mass of new facts became known to us. We found that in starvation protein and fat alone were burned, that during work more fat was burned and that less

fat was consumed during rest, especially during sleep; that the carnivorous dog could maintain himself upon an exclusive protein diet, and if to such a protein diet fat were added, the fat was almost always entirely deposited in the body; that carbohydrates, on the contrary, were burned no matter how much was given and that they, like the fat of the food, protected the body from fat loss, although more carbohydrates than fat had to be given to effect this purpose; that the metabolism in the body was not proportional to the combustibility of the substances outside the body, but the protein which burns with difficulty outside, metabolizes with the greatest ease, then carbohydrates, while fat, which readily burns outside, is the most difficult to combust in the organism (Kinney 1988: 524).

Max Rubner (1854–1932), who was trained in the laboratories of Voit, determined standard food energy (caloric) values for the major foodstuffs. He continued to demonstrate, until well into the twentieth century, that the animal body followed the law of the conservation of energy.

W. O. Atwater (1844–1907), an American who had studied under Voit in Germany, returned to the United States to work with a physicist at Wesleyan University in developing a large calorimeter capable of measuring precisely the amount of heat given off by a person living in it.¹ This calorimeter confirmed the earlier experimentation of Voit, Pettenkofer, and Rubner and demonstrated that the energy expended by a person in doing any work, such as static bicycle riding, was equal to the heat released by the metabolism of food in the body. The physiologist F. G. Benedict extended Atwater's work by constructing special equipment at the Carnegie Institute in Boston for the simultaneous measurement of gaseous metabolism and heat production. An example from one of their studies is shown in Table IV.C.3.1. On a daily basis, the difference between heat determined and heat estimated was only –1.6 percent. The precision obtained in these experiments was remarkable and was obtained without computer assistance!

The finding from the calorimeter work that received the most attention in Atwater's lifetime concerned alcohol. In 1899 he reported that if a subject drank alcohol in small portions over the course of the day, it was almost fully oxidized and replaced the food energy equivalent of either fat or carbohydrate. At this level, therefore, it acted as a food (Atwater and Benedict 1899). This finding was advertised by the liquor trade, and Atwater was subsequently attacked by temperance advocates as a disgrace to his church and to his profession (Pauly 1990). He defended himself vigorously and, while agreeing that abstinence was certainly the safest way to avoid addiction, argued nevertheless that for most people,

Table IV.C.3.1. *Energy exchange in humans: An example from Atwater and Benedict*

	Total: 4 days	Daily average
a. Heat of combustion of food eaten	9848	2462
b. Heat of combustion of feces	304	76
c. Heat of combustion of urine	538	135
m. Heat of combustion of alcohol eliminated	84	21
d. Estimated heat of combustion of protein gained (+) or lost (–)	–277	–69
e. Estimated heat of combustion of fat gained (+) or lost (–)	–539	–135
f. Estimated energy of material oxidized in the body a – (b + c + m + d + e)	9738	2434
g. Heat determined	9576	2394
h. Heat determined greater (+) or less (–) than estimated (f – g)	–162	–40
i. Heat determined greater (+) or less (–) than estimated (h + f) (percent)	–	–1.6

Source: Atwater and Benedict (1899).

moderate consumption was possible and seemed to improve their well-being rather than damage it (Carpenter 1994b). (We have heard recent echoes of this debate following recommendations in the Dietary Guidelines for the United States concerning alcohol consumption.)

Further related studies undertaken by Atwater (1899) and his associates (Atwater and Bryant 1900), which overlapped those of Rubner, made determinations of the fuel value of foods by the use of a bomb calorimeter and by comparing those determinations with the heat produced in the body. A bomb calorimeter permits the complete combustion of the sample in the presence of oxygen under pressure, with the direct determination of the heat produced by measurement of the temperature rise in the surrounding water. A diagram of Atwater's bomb calorimeter is shown in Figure IV.C.3.1.

It was found that the total (gross) energy produced by combustion of food with oxygen in a bomb calorimeter was greater than the metabolizable energy of the same food when utilized in the body. However, when account was taken of digestibility (that is, not all the food consumed was absorbed into the body) and of the energy lost in the urine (that is, the urea produced from protein), then there was essential agreement. Fuel values of foods were compared with their analytical composition, and thus were born the 4:9:4 Atwater factors, where the metabolizable energy of foods, in kilocalories per gram, could be readily calculated from the contents of carbohydrate, fat, and protein. These concepts were reported, essentially in their present form, in the texts of R. Hutchison (1903) and H. C. Sherman (1914).

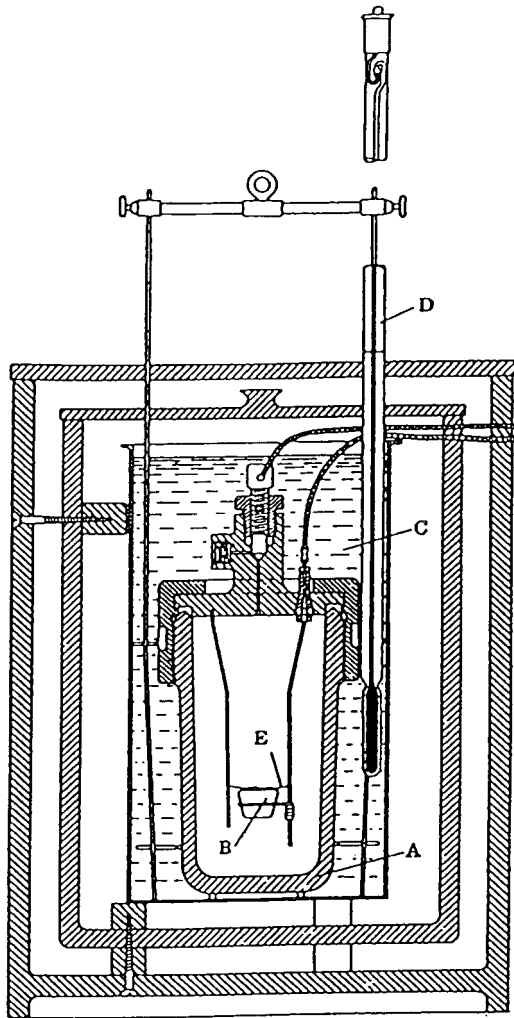


Figure IV.C.3.1. The Atwater bomb calorimeter: (A) heavy steel bomb filled with oxygen under pressure; (B) capsule inside bomb containing weighted sample; (C) weighed amount of water at known initial temperature; (D) thermometer of high accuracy; (E) electric ignition wire. (From Sherman 1952.)

The unit used today for food energy, deriving from these and other basic physical studies, is the kilocalorie (kcal). This is the amount of heat necessary to raise 1 kilogram of water from 15° to 16° Celsius. The international unit of energy is the more scientifically fundamental joule (J). Although not greatly used in the United States in a nutritional context, the joule is widely employed in Europe. To convert from kilocalories to kilojoules (kJ), the rounded value 4.2 may be used (1 kcal = 4.184 kJ). The classical Atwater metabolizable energy conversion factors of 4 kcal/gram (g) of food protein and carbohydrate and 9 kcal/g of food fat have been verified and are adequate for computa-

tion of the energy content of customary diets (Miles, Webb, and Bodwell 1986). Alcohol (ethanol) has a computed food energy value of 7 kcal/g or 5,600 kcal/liter (L). It should be noted that in the original Atwater calculations, carbohydrate was measured by difference (that is, only protein, fat, and minerals were determined in the food, and the residue after allowing for water content was considered to be mainly carbohydrate). This, therefore, included fiber.

The Development of Modern Nutritional Science

Although the complex details of intermediary metabolism remained to be discovered as the twentieth century began, the basic framework of the modern science of nutrition had been laid down. Free sugars and their polymers furnished glucose for bodily functions, including muscular work: Excess carbohydrate could be stored as glycogen and fats, thus providing a fuel reserve that could furnish energy during periods of fasting. The views of Liebig concerning heat and work were recognized as mistaken by Hutchison in his text of 1903:

We now know that bodily heat is not a thing apart and requiring to be provided for by itself, but that it is an inevitable accompaniment of cell life. Life and heat are inseparable, and in fulfilling its other function in the body a cell cannot help producing heat also (Hutchison 1903:3).

Hutchison went on to indicate that as far as the cells of the body were concerned, it was a matter of indifference whether carbohydrate, protein, or fat were the original sources of the food energy. In strict energy terms this remains true, although the specific role of the energy sources, especially fat, in relation to chronic disease risk is now central to dietary guidance.

Lusk, author of several monographs and texts on nutrition (Lusk 1914, 1928, and 1933), and yet another who trained with Voit in Germany, dedicated the first edition of his monograph *The Elements of the Science of Nutrition* to his mentor in 1906. That an essentially modern view of nutrition had developed by this time can be seen from Lusk's writing (1914:4):

The workshops of life require fuel to maintain them and a necessary function of nutrition is to furnish fuel to the organism so that the motions of life continue. Furthermore the workshops of life are in a constant state of partial breaking down and materials must be furnished to repair worn out parts. In the fuel factor and the repair factor lie the essence of the science of nutrition.

Lusk (1914) also held modern views on the social aspects of nutrition, as is indicated in his lecture to the New York Academy of Medicine, which advocated

nutritious school meals, family nutrition education, and national nutritional labeling of foods.

Another major activity in the early years of the twentieth century became the determination of the basal metabolic rate (BMR) in many subjects together with the development of prediction procedures for estimating BMR from body weight or from body surface area (Harris and Benedict 1919). The BMR is the rate of heat production, measured directly or indirectly, when the subject is at rest and has not eaten for several hours. The normal range for large populations was thought to be relatively narrow. Variations in the BMR of supposedly normal individuals were often attributed to carelessness in experimental procedure, rather than to the existence of real differences. The field of calorimetry became devoted mainly to examining abnormalities in BMR associated with various diseases, in particular, thyroid disease, diabetes, and the more common fevers.

Equipment for determining BMR was standard in hospital laboratories until about 1950, at which time new chemical methods for measuring materials in the urine and blood related to thyroid metabolism caused hospitals to abandon the measurement and, consequently, to terminate the last remaining quantitative approach to energy metabolism in patients. In the hospital environment, research on energy expenditure in surgical patients returned under the leadership of Kinney (1988), who utilized a head canopy system that was acceptable to acutely ill patients and still permitted medical care as required.

There has been an enormous resurgence of interest in energy balance studies in recent years as a result of the growing concern over the hazards of obesity and the realization that we still have much to learn about energy metabolism and the control of energy balance.

World Supply of Food Energy

The energy, fat, and protein supply in rich and poor countries is shown in Table IV.C.3.2. Carbohydrate is the major source of food energy in both industrialized (52 percent) and developing (70 percent) countries. For the rich industrialized countries, the fat proportion (35 percent) is next in importance, but this proportion is much lower (20 percent) in developing countries. Whereas the proportion of the food energy as protein (PCal percent) is similar (12 percent to 10 percent) in both sets of countries, the proportion of that protein originating from animal sources differs greatly and is very much lower in developing countries, whose people receive considerably more of their protein from cereals.

The main dietary carbohydrates are shown in Table IV.C.3.3. Carbohydrates produce energy for metabolism through their initial conversion to acetyl coenzyme A (acetyl CoA), which is also the central point in the metabolism of protein and lipids for

Table IV.C.3.2. *Food availability data for industrialized and developing countries. Data for 1994*

	Industrialized countries	Developing countries
Food energy kcal/day	3356	2573
Total protein g/day	103.4	64.4
Animal protein g/day	62.2	17.1
Fat g/day	131.9	56.1
Derived data		
PCal% ¹	12.3	10.0
FCal% ²	35.4	19.6
Animal protein % total protein	60.2	26.6
Cereal protein % total protein	24.3	53.3
Sugar as % total calories	9.5	5.9
Carbohydrates as % total calories	52.3	70.4

¹Protein calories as percentage of total calories.

²Fat calories as percentage of total calories.

Source: FAOSTAT (1996).

Table IV.C.3.3. *The principal dietary carbohydrates*

Source	Maize, rice, wheat, potato	Cane and beet sugar	Milk
Polysaccharide	Starch	—	—
Disaccharide	Maltose	Sucrose	Lactose
Monosaccharide	Glucose	Fructose + Glucose	Galactose + Glucose

energy (Figure IV.C.3.2). Because storage fats can be synthesized from acetyl CoA, there is some degree of interconversion of the macronutrients, and we have the biochemical explanation of the fat synthesis controversy that taxed the early physiological chemists. The reverse conversion cannot occur to any significant degree, and fat is not a source of carbohydrate and protein.

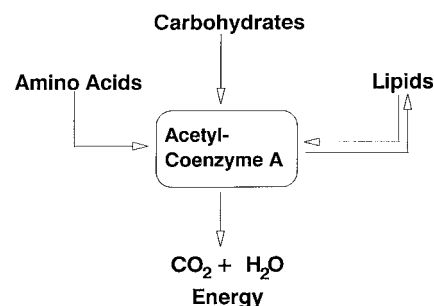


Figure IV.C.3.2. An overview of the combustion of fuels for energy. Acetyl coenzyme A is formed from carbohydrates, fats, and amino acids. Whatever the source, a large fraction of it is oxidized to CO₂ and H₂O, although any excess may be used to form lipids as an energy reserve.

Lipid is the overall chemical term used to describe the fats and oils; of these, the triglycerides (glycerol esterified with 3 fatty acid molecules) are the most abundant form in the diet. Although lipids differ widely in their compositions, the major pathways for their synthesis and breakdown begin and end with acetyl coenzyme A. Proteins are converted to acetyl CoA through a number of other intermediaries common mainly to the metabolism of carbohydrates. An outline of these latter pathways is shown in the section “Amino Acid Metabolism.”

Thus, the energy of food is liberated by chemical changes, mostly oxidations, and is converted into mechanical work and heat. It is used to power the organic syntheses necessary for tissue maintenance and for growth in the young, and energy is also needed for the internal work of the body, the beating of the heart, the movements of the respiratory muscles, and the maintenance of the differences in the chemical contents of the cell fluids and the body fluids. Present descriptions of these processes include such terms as “biochemical energetics” and “free energy,” but these modern explanations still conform to the broad descriptive strokes used in the early years of the century. The concept of burning fuels, not only for the energy with which to move but also for the energy necessary to build tissues and the like, is our rationalization of these processes (McGilvery 1979). Examples of the more important energy-producing and energy-consuming reactions in plants and animals include the following:

Energy-producing (exergonic)	Energy-consuming (endergonic)
Oxidation of fuels (carbohydrates, fats, and proteins)	Mechanical movement
Photosynthesis	Synthesis of cellular constituents
Fermentations	Creation of concentration gradients
	Storage of fuels

The sequence of reactions in the cell by which acetyl groups are oxidized is the “citric acid cycle” (or “Krebs cycle”), and for this discovery made in the 1930s, Hans Krebs (1900–81) shared the Nobel Prize in 1953. The reactions of the cycle function in sequence and as a unit to remove electrons from acetyl groups within the mitochondria and to feed the electrons into the oxidative phosphorylation pathway with the ultimate transfer to molecular oxygen. The overall result is the production of energy-rich adenosine triphosphate (ATP) from the oxidation of acetyl groups. A simple overview of energy production from the macronutrients is illustrated in Figure IV.C.3.2.

Human Energy Expenditure and Its Components

The main factors that influence daily energy need are basal metabolic rate (BMR) or resting energy expenditure (REE), physical activity, and thermogenesis. The

last term may produce confusion for nonspecialists because the word means the generation of heat, and all energy expenditure generates heat. Thermogenesis is additional heat generation above that of REE and physical activity. All factors are affected directly or indirectly by individual variables, such as age, sex, body size, and, to a lesser degree, climate.

Resting Energy Expenditure

Unless the activity level is high, the most important single component of overall energy requirement is the energy expended at rest (REE). Customarily, basal metabolism, or basal metabolic rate (BMR), has been included as one of the factors affecting overall energy requirement. In practice, however, the chief interest is in resting energy expenditure (REE), namely, the energy expended by a person in a normal life situation while at rest and under conditions of thermal neutrality. The REE includes the thermic effect of meals and is an average minimal expenditure for the periods of the day and night when there is no exercise and no exposure to cold. REE is influenced by, among other things, age, height, weight, and sex. Although no equation for predicting REE can be completely accurate, several include age, sex, weight, and height and can predict REE with sufficient accuracy to serve as a first step in determining group or individual requirements.

An animal loses heat through the surface of the body, and the capacity for heat production is related to the volume of the body. It was therefore reasoned in the early 1800s that the larger the animal, the greater the heat production relative to the heat loss. Rubner later proposed that when metabolic rate was expressed in relation to the body surface area, the effect of body size disappeared. More elaborate studies were undertaken in the 1930s by Kleiber (see Kleiber 1961), who presented strong evidence against surface area being the determining factor of basal metabolism. Rather, he suggested that metabolism should be considered as a power function of body weight. Lacking a better definition of metabolic size, he proposed a three-fourths power of body weight as the best correlation between body size and resting metabolism. Later relationships were proposed with fat-free body mass, rather than with body weight alone. Lean body mass is a more recent term, and the concept has been further extended to the metabolizing body-cell mass by measuring total exchangeable potassium (Kinney 1988). Nevertheless, for humans, weight is the major variable and may, together with age and sex, be sufficient for predictive purposes.

The Food and Agriculture Organization of the United Nations (FAO) equations (FAO/WHO/UNU 1985) are shown in Table IV.C.3.4. The Harris-Benedict equations, which also include height, have been widely used in the United States for predicting the energy requirements of hospitalized patients; they are also shown in Table IV.C.3.4. In relation to the Harris-

Benedict equations, Sherman (1952) referred to their empirical nature and remarked that they “may be somewhat disconcerting at first glance. On working through the two formulas, however, it will be seen that there is . . . [no] great difference in the basal metabolisms of the sexes” (Sherman 1952: 161).

Thermogenesis

The thermic responses to food and to cold were formerly thought to be independent. N. J. Rothwell and M. J. Stock (1983) proposed that brown adipose tissue may perform a regulatory function in response to overfeeding. Rats could become obese on a palatable snack-food diet but maintained normal body weights when given less palatable rat chow. The brown adipose tissue of the rats with abnormal weight gain was markedly increased, suggesting that the overfeeding may have led to a substantial increase in dietary induced thermogenesis.

Diet-induced thermogenesis (DIT) is any change in energy expenditure induced by diet. This can be divided into obligatory and adaptive components. DIT reaches a maximum about 1 hour after the meal is

consumed and virtually disappears 4 hours afterward. There have been reports of subjects who have apparently eaten excessively for long periods but who have gained much less weight than would have been predicted had they stored this excess as adipose tissue. The efficiency with which energy intake is utilized may vary greatly within the same individual and may explain the large variations noted in many experiments. Variations in thermogenesis have been an important research area in explaining some of the discrepancies between theory and actual observations of energy balance. Some researchers, however, now claim that the energy intake data in these observations may have been underestimated and that there may be no significant discrepancies (Heymsfield et al. 1995; Seale 1995).

Physical Activity

A dominant factor influencing overall energy needs is the proportion of time an individual devotes to moderate and heavy activities, in contrast to those that are light or sedentary. Little is known, however, about why apparently similar individuals differ in their selected

Table IV.C.3.4. *Equations for predicting basal metabolic rate from body weight (W) and age*

Equations from FAO				
Age range (years)	kcal/day	Correlation coefficient	MJ/day	Correlation coefficient
Males				
0-3	60.9 W - 54	0.97	0.255 W - 0.226	0.97
3-10	22.7 W + 495	0.86	0.0949 W + 2.07	0.86
10-18	17.5 W + 651	0.90	0.0732 W + 2.72	0.90
18-30	15.3 W + 679	0.65	0.0640 W + 2.84	0.65
30-60	11.6 W + 879	0.60	0.0584 W + 3.67	0.60
>60	13.5 W + 487	0.79	0.0565 W + 2.04	0.79
Females				
0-3	61.0 W - 51	0.97	0.255 W - 0.214	0.97
3-10	22.5 W + 499	0.85	0.0941 W + 2.09	0.85
10-18	12.2 W + 746	0.75	0.0510 W + 3.12	0.75
18-30	14.7 W + 496	0.72	0.0615 W + 2.08	0.72
30-60	8.7 W + 829	0.70	0.0364 W + 3.47	0.70
>60	10.5 W + 596	0.74	0.0439 W + 2.49	0.74

W = weight in kilograms.

Source: FAO/WHO/UNU (1985).

Harris-Benedict equations

For males: $H = 66.473 + 13.752 W + 5.003 S - 6.755 A$

For women: $H = 65.501 + 9.563 W + 1.850 S - 4.676 A$

H = Total heat production in 24 hours

W = Weight in kilograms

S = Stature in centimeters

A = Age in years

Source: Harris and Benedict (1919).

level of physical activity (Hill et al. 1995). Average physical activity in Western society is now considered to be undesirably low. Estimates of energy needs were originally based on physical activity associated with different occupations, but occupational needs have declined significantly in recent years with the introduction of labor-saving machinery and devices. Simultaneously, there has been an increase in many discretionary activities, resulting in great variability among individuals. The widespread adoption of jogging and other forms of fitness activities are obvious examples. Thus, the traditional specification of energy needs based on occupation is no longer adequate.

Three categories of discretionary activity that affect energy needs have been recognized (FAO/WHO/UNU 1985): optional household activities, which include working in the garden or repairing and improving the home; socially desirable activities, which are especially important contributors to the energy needs of children, who require additional energy to explore their surroundings, to learn, and to adjust their behavior in order to relate to other children; and physical fitness activities, which contribute to the well-being of sedentary workers through leisure-time exercise. It is now considered more important in weight control to recommend increased physical activity, rather than reduced energy intake, because there is a range of benefits to health beyond the maintenance of energy balance (Conway 1995). This is in sharp contrast to the recommendations made by the American Medical Association in 1951 whereby it was stated that “it is seldom wise to attempt the alteration of calorie balance by changing the energy output; it is far easier and surer to adjust the intake” (AMA 1951:274).

Other Factors Affecting Energy Requirements

Age. REE varies with the amount of metabolically active tissue in the body, with the relative proportion of each tissue in the body, and with the contribution of each tissue to total metabolism. These all vary with age. Changes in the body composition of children, adults, and elderly people are taken into account when energy requirements of a particular population group are calculated by use of the empirical equations that link REE with age. Activity patterns also change with age. Children become progressively more active once they are able to walk, and the physical activity patterns of adults are dependent not only upon their occupation but also upon their own leisure activity. When adults retire from work, their change in habits must be recognized when estimating energy requirements.

Sex. Although there are differences in the body weight and body composition of boys and girls beginning during the first few months of life, such differences are relatively small until children reach 9 to 10 years of age. Changes then accelerate throughout ado-

lescence. After maturity, men have greater muscle mass, whereas women's bodies generally contain a greater proportion of fat. In practice, REE per unit of total body weight is not greatly different in men and in women. The energy requirement for physical activity has often been related to different occupational needs. In recent years, however, occupational and discretionary activity requirements of both sexes have become increasingly similar.

Body size. Persons with large (or small) bodies require proportionately more (or less) total energy per unit of time for activities such as walking, which involve moving mass over distance. Their hourly rates of REE will also be slightly higher or lower than the average. Energy allowances must be adjusted for the variations in requirements that result from such differences in body size. Weight may be used as a basis for adjusting allowances for different body sizes within a given age and sex category, provided the individuals are not appreciably over or under their desirable weights. Persons who are overweight often compensate for the increased energy cost of carrying their extra weight by decreasing daily activity.

Climate. In industrialized societies, most persons are protected against cold by warm clothes and heated environments. Many also live and work in air-conditioned buildings, so the effects of high temperatures are reduced. Yet, everyone is not insulated from environmental exposure, and when there is prolonged exposure to cold or heat, energy allowances may need adjustment.

Food Energy Requirements

The energy requirement of an individual has been defined as

[t]hat level of energy intake from food which will balance energy expenditure when the individual has a body size and composition, and level of physical activity, consistent with long-term good health; and which will allow for the maintenance of economically necessary and socially desirable physical activity. In children and pregnant or lactating women the energy requirement includes the energy needs associated with the deposition of tissues or the secretion of milk at rates consistent with good health (FAO/WHO/UNU 1985: 12).

As the definition suggests, energy expenditure can often be the basis for estimating energy requirements. The most frequently used procedure for determining energy expenditure consists of a combination of a timed activity record (that is, the total duration of each of the important activities throughout the whole 24 hours of the day) and an energy value in kcal or kJ per minute for each of these activities. These energy

values may be derived from either published data or from actual measurement of oxygen consumption. Time spent at BMR or REE (sleep or rest) can often be a major component of the day's need.

A modified procedure for estimating average daily expenditure is to multiply the basal metabolic rate (BMR) or resting energy expenditure (REE) by an appropriate daily *average* activity factor, which is dependent on the degree and duration of physical activity. The latter is often termed the PAL (physical activity level). Thus: Daily Energy Expenditure = BMR $\times$ PAL. The PAL may vary from about 1.2 to 1.4 for relatively inactive people but up to 2.0 or more in the case of people who are physically very active. The value of the daily PAL may be obtained from questionnaires on habitual physical activity. Estimated values (FAO/WHO/UNU 1985; IDECG 1996) for PALs at various levels of habitual activity at different ages are shown in Table IV.C.3.5.

The doubly labeled water (DLW) method (Lifson, Gordon, and McLintock 1955) for the measurement of energy expenditure makes use of the stable isotopes ^{18}O and ^2H . The rates of disappearance from body water are determined following the oral administration of a known amount of doubly labeled water. Both energy expenditure and water output over several days can subsequently be calculated. Although the theoretical advantages of the DLW procedure, such as the ability to provide data on free-living individuals in almost any context, are considerable, there are counterbalancing drawbacks (Durnin 1996): The technicalities of the analysis of CO_2 output are such that expensive equipment and technical expertise are highly necessary. In addition, the isotope ^{18}O is expensive. Lastly, the calculations only provide information on CO_2 output for a period of several days. Nevertheless, the procedure is now widely used (Conway 1995), and many of the requirement data for various population groups – as well as estimates of PALs in the IDECG (1996) “Energy and Protein Requirements” report – are

from DLW determinations. Comparisons between DLW and other procedures, demonstrating a close agreement in findings, have been reported by Seale (1995).

Respiration chambers or whole-body calorimeters can have a role in assessing requirements but are highly expensive and technically complex. Few of these chambers exist, and their usefulness is restricted to specific basic problems that do not require a natural, free-living environment. But such devices can still play a research role in studying the relationship between energy metabolism and heart rate, diet-induced thermogenesis, DLW validation, and the influence of varying proportions of energy-supplying nutrients on energy metabolism. Extrapolation from heart rate to energy expenditure is useful for certain population groups, such as young children, the elderly, and those who are ill.

Energy expenditure has sometimes been indirectly estimated from energy intake, although the potential for error can be enormous, especially when intake is estimated from an interview based on a remembered food consumption for one day (24-hour recall). Self-recording of energy intake also frequently underestimates intakes when compared with measurements made by more direct procedures (Heymsfield et al. 1995; Seale 1995). However, when the procedure for measuring energy intake has been performed in an acceptable fashion by experienced observers, the results may be essentially the same as energy expenditure simultaneously measured (Durnin 1996).

Despite a much greater understanding of energy balance, many discrepancies still exist between self-reported food energy intakes and weight gain. Comparisons have been made by S. B. Heymsfield and colleagues (1995) between actual (DLW and indirect calorimetry) and estimated (self-reported) energy intakes. The studies revealed a considerable degree of underestimation of energy intake when using self-reporting procedures. In consequence, many of the cited discrepancies between food energy intake and weight gain may not really exist, and thermogenesis may not need to be invoked to explain the differences.

Table IV.C.3.6 presents energy intakes and requirements for various population groups in the early 1900s. The high requirement values it reveals were reasonable and acceptable when much manual labor was essential. However, the more recent international recommendations shown in Table IV.C.3.7 are much lower for all age groups. In Table IV.C.3.8, data on food energy and protein availability for a number of world regions are presented for the year 1994. These data are population averages, indicating that many people in the least developed countries are likely to be at risk of food energy inadequacy, whereas many of those in developed regions may be suffering from surfeit. The differences in protein availability, especially that from animal sources, are also noteworthy.

Table IV.C.3.5. *Physical activity levels suggested to estimate total daily energy expenditure from the mean basal metabolic rate of children, adolescents, and adults*

Age (y)	Sex	Habitual physical activity		
		Light	Moderate	Heavy
1–5	M, F	1.45	1.60	–
6–13	M	1.55	1.75	1.95
14–18	M	1.60	1.80	2.05
6–13	F	1.50	1.70	1.90
14–18	F	1.45	1.65	1.85
Adult	M	1.55	1.80	2.10
	F	1.55	1.65	1.80

Source: Child and adolescent data IDECG (1996). Adult data FAO/WHO/UNU (1985).

Table IV.C.3.6. *Early protein and energy intakes from Europe and America with requirement estimates*

Intake and requirement groups	Protein gm/day	Fat gm/day	kcal	PCal%	FCal%
Europe					
Blacksmith U.K.	176	71	4117	17.1	15.5
Brickmaker Germany	167	117	4641	14.4	22.7
Carpenter Germany	131	68	3194	16.4	19.2
Tailor U.K.	131	39	3053	17.2	11.5
Workman Russia	132	80	3675	14.4	19.6
Workman Sweden	134	79	3436	15.6	20.7
North America					
Factory worker Canada	109	109	3622	12.0	27.1
Glassblower Massachusetts	95	132	3590	10.6	33.1
Mechanic Massachusetts	127	186	4428	11.5	37.8
U.S. Army ration	120	161	3851	12.5	37.6
U.S. Navy ration	143	184	4998	11.4	33.1
U.S. average 53 students	103	138	3500	11.8	35.5
Requirements for hard work					
Voit	145	100	3370	17.2	26.7
Rubner	165	70	3644	18.1	17.3
Playfair	185	71	3750	19.7	17.0
Atwater	150	150	4060	14.8	33.3
Diets for normal work					
Atwater	112	148	2992	15.0	44.5
Chittenden	58	140	2729	8.5	46.2
Chittenden low intake	41	79	1584	10.4	45.1

PCal%: Protein calories as percentage of total calories.

FCal%: Fat calories as percentage of total calories.

Source: Hutchinson (1903). Normal work requirements data for Atwater and Chittenden from Sherman (1914).

Table IV.C.3.7. *International food energy requirements 1950–96*

	Weight kg	FAO 1950	FAO/WHO/UNU 1985 kcal/kg/d	IDECG 1996
0–6 mo.	3.6–7.6	–	118	90
6–12 mo.	8.0–9.7	110	94	88
Children and adolescents		kcal/d		
1–3 yr.	12.2	1200	1380	922
4–6	18.2	1600	1750	1323
7–9	25.3	2000	2030	1587
10–12	36.3	2500	2068	1894
13–15 F	49.9	2600	2140	2007
13–15 M	50.9	3200	2515	2457
16–20 F	56.3	2400	2135	2098 ¹
16–20 M	65.5	3800	2835	2908 ¹
Adult F	55.0	2300	2100	2100
Adult M	65.0	3200	2700	2700
Pregnancy	–	+ 450	+ 285	+ 225
Lactation	–	+ 1000	+ 500	+ 685

¹16–18 years.

Note: Adult requirements are for light-to-moderate activity levels.

Some international recommendations for requirements of food energy, together with discussions of the issues involved, have been provided by FAO (1950), FAO/WHO/UNU (1985), and IDECG (1996), whereas recommendations for the United States are discussed by Pellett (1990a). Food-energy requirements specifically for adults are dealt with by Shetty and colleagues (1996), and pregnancy and lactation are considered by Prentice and colleagues (1996).

Protein

Historical Background

Proteins are associated with all forms of life, an observation that dates back at least to the original identification of proteins as a class by G. J. Mulder (1802–80), who wrote in 1840:

In both plants and animals a substance is contained, which is produced within the former, and imparted through their food to the latter. It is unquestionably the most important of all known substances in the organic kingdom. Without it no life appears possible on this

Table IV.C.3.8. *Distribution of food energy, fat, and protein in the various world regions*

Data for 1994	Food energy kcal/d	Total protein g/d	Animal protein g/d	Plant protein %	Animal protein %	Cereal protein %	FCal%
World	2718	72.1	25.8	64.2	35.8	45.9	23.1
Developed countries	3206	98.0	55.2	43.7	56.3	29.6	32.5
Industrialized countries	3356	103.4	62.2	39.8	60.2	24.3	35.4
North America, developed	3591	112.6	72.1	35.9	64.0	21.9	35.3
Oceania, developed	3123	102.2	70.7	30.8	69.2	18.7	35.3
Western Europe	3398	102.5	61.5	40.0	60.0	24.3	39.5
Developing countries	2573	64.4	17.1	73.6	26.6	53.3	19.6
Africa, developing	2336	58.3	11.4	80.4	19.6	53.9	18.1
Asia, developing	2600	64.8	16.0	75.3	24.7	56.0	19.0
Latin America, developing	2732	70.4	31.9	54.7	45.3	35.7	25.4
Least developed	2013	49.9	9.0	82.0	18.0	58.9	14.8

FCal% = Fat calories as percentage of total calories.

Source: Derived from FAOSTAT (1996).

planet. Through its means the chief phenomena of life are produced. (Block and Bolling 1951: v)

We now recognize that the major importance of protein in the diet is to act as a source of amino acids, some of which are essential (indispensable) dietary constituents. This is because their carbon skeletons are not synthesized in the bodies of animals. Others are nonessential (dispensable) because they can be made within the animal from readily available carbon and nitrogen precursors. All animal species need some pre-formed amino acids in their diets (Munro and Crim 1988). The essential amino acids (EAAs) for humans include histidine, isoleucine, leucine, lysine, methionine, phenylalanine, threonine, tryptophan, and valine; moreover, cysteine and tyrosine are synthesized in the body from methionine and phenylalanine, respectively.

An additional 9 amino acids (alanine, arginine, aspartic acid, asparagine, glutamic acid, glutamine, glycine, proline, and serine) are also present in proteins but are not necessary in the diet because the body can usually manufacture them as needed, provided there is an adequate source of nitrogen and glucose. Sufficient synthesis may not take place in some circumstances, and some nonessential amino acids may become conditionally essential. This can occur, for example, in preterm infants (Tsang et al. 1993), following liver damage in the adult, and in the presence of certain potentially toxic materials.

The proteins of living matter act, among other things, as enzymes, as structural features of the cell, as messengers (peptide hormones), and as antibodies. The accumulation of proteins during growth and development, and the maintenance of tissue proteins in the adult, represent important objectives in ensuring nutritional well-being. A knowledge of the way in which dietary protein is optimally utilized by the body allows determination of protein needs for health and for the restoration of body tissue following disease (Munro and Crim 1988).

Amino Acids and the Quality of Protein

Before the first decade or so of the twentieth century, it was thought that proteins were all of generally similar nutritional value, despite recognition of very obvious differences in their chemical and physical properties.

Rubner, however, in the late 1890s, observed that proteins of varying origins were not of the same value in nutrition, and the earlier work of the Magendie Commission on Gelatin had demonstrated that gelatin could not support life (McCollum 1957). But these efforts and numerous nitrogen balance studies performed after the 1850s were essentially disregarded until the beginning of the next century (see Munro 1964a) when work reported from England confirmed that the different values resulted from lack of amino acids (Willcock and Hopkins 1906–7).

Sherman (1914), in the first edition of his textbook (the eighth edition appeared in 1952), defined protein in basically modern terms: "The word protein should designate that group of substances, which consist, as far as at present is known, essentially of combinations of alpha-amino acids and their derivatives" (1914: 26). By 1914, Lusk, in a lecture to the New York Academy of Sciences on the importance of proteins, could provide an almost complete list of those amino acids recognized today as constituents of proteins, with methionine and threonine the most notable absentees.

Within a decade of Lusk's lecture, Thomas B. Osborne and Lafayette B. Mendel at Yale University had convincingly demonstrated that there were major differences in the quality of proteins and had established the basis of our present evaluation techniques. Mendel suspected that the concentration of protein in a diet was not a completely satisfactory indicator of its ability to support growth, whereas Osborne had isolated a number of pure proteins. Thus began a most fruitful collaboration whereby Mendel studied

the growth response of rats that were fed the proteins isolated by Osborne.

They found, quite early in their studies, that the various isolated proteins differed in their potential to support growth because the weight gain (of rats) per unit of protein consumed varied significantly. Hence was born the PER or protein efficiency ratio (Osborne, Mendel, and Ferry 1919), which remained the official method for the evaluation of protein in the United States (Pellett 1963) until the protein quality evaluation report by the FAO/WHO in 1991. From these collaborative studies came the early conclusions that rats grew better on animal proteins than on plant proteins, but that adding some of the then newly isolated and chemically identified amino acids to plant proteins could make them equal to animal proteins in their growth-promoting ability. The broad principles of protein nutrition that we accept today were well known by the second decade of the twentieth century (Mendel 1923).

Whether, however, there were fundamental differences between animal and vegetable proteins remained an important issue. It was a debate in which agreement was reached much earlier in scientific circles than in the public arena, where it remained an issue at least until the 1950s (Block and Bolling 1951). By that time it was agreed that an essential amino acid from a plant was the same as one from an animal but that the proportions present in plant and animal foods could differ and that mixing proteins in the diet (complementation) could confer nutritional benefits.

Nitrogen balance techniques had been in use at least since Boussingault in 1839, but the determination of the biological values of proteins from nitrogen balance data is generally attributed to K. Thomas (1909). The ability to scale the nutritional differences in proteins numerically arose in the first and second decades of the twentieth century with the introduction of the PER by Osborne, Mendel and E. L. Ferry (1919) and the refinement of Thomas's nitrogen balance techniques by H. H. Mitchell in 1924.

Protein and Amino Acid Analysis

Before nitrogen balance and other biological methods for the evaluation of protein quality could be used in any routine manner, improvements in procedures for the analysis of protein were required. Although the Dumas procedure (which converted the nitrogen of protein to gaseous nitrogen that could then be measured) had been available since 1835, it was not until 1883 that J. Kjeldahl described a convenient procedure for converting the nitrogen of most organic materials into ammonia. The procedure involved boiling the material in the presence of concentrated sulfuric acid and a catalyst. After release with a strong base, the ammonia could be determined by acid-base titration. Numerous modifications of the method have

been proposed, but it remains remarkably similar in its essentials to the original procedure and is still widely used.

Nitrogen in foods comes not only from the amino acids in protein but also from additional forms that may, or may not, be used as part of the total nitrogen economy in humans and animals. The nitrogen content of proteins in foods can vary from between about 15 to 18 percent, depending on the amino acids in these proteins. In addition, purines, pyrimidines, free amino acids, vitamins, creatine, creatinine, and amino sugars all contribute to the total nitrogen present. Urea is also a major contributor in foods as important as human milk.

Because the average content of nitrogen (N) in protein is 16 percent, the nitrogen content multiplied by 6.25 is frequently used to convert N to protein. This should then be termed "crude protein." It is not always appreciated that the amino acids themselves show a very large range in their content of nitrogen. Whereas tyrosine has less than 8 percent N, lysine has 19 percent and arginine contains 32 percent. The presence of an amide group in an amino acid can double the N content; aspartic acid contains 10 percent N, but asparagine contains 21 percent. For other N-containing compounds, values can even be higher, with 47 percent N in urea and some 80 percent in ammonia. In practice, most protein evaluation techniques are actually evaluating the nitrogen present but are calling it protein ($N \times 6.25$). Because a number of other factors from 5.3 to 6.38 (Jones 1941) may be used to convert food nitrogen to protein in different foods, considerable potential for confusion exists (Pellett and Young 1980).

Early in the twentieth century as the ability developed to detect and identify amino acids, it was realized that they were present in living materials, not only as components of their proteins but also as free amino acids. It also became clear that there were many more amino acids in living systems than the 20 or so known to be present in proteins. Because all amino acids behave in a rather similar chemical manner, additional techniques, often of a physicochemical nature, such as column chromatography, became necessary to solve the complex problems of separating and determining their presence in foods (Moore and Stein 1954; Moore, Spackman, and Stein 1958).

A major milestone was the description (Spackman, Stein, and Moore 1958) of an automatic analyzer allowing separation and quantitative analysis of amino acids based on ion-exchange column chromatography. As a result, modern dedicated amino acid analyzers quantitatively separate protein hydrolysates into their constituent amino acids in 1 hour or less and physiological fluids in about 2 hours.

In the introduction to the first edition of their major work on the amino acid composition of food proteins, R. J. Block and D. Bolling wrote:

It has been our experience that a food protein may be a good source of those nutritionally valuable amino acids which are most commonly estimated (i.e., cystine, methionine, arginine, histidine, lysine, tyrosine and tryptophan) and yet be deficient in one or more of the other essentials for which analytical methods are more difficult and often less accurate (1951: xiv).

Threonine was not listed, but the other essential amino acids considered at the time to be difficult to analyze (leucine, isoleucine, valine, and phenylalanine) have rarely proved to be in deficit in normal foods or diets. Analytical techniques for amino acids at the time were chemical and followed in a direct manner from those developed by Fischer (1901). Large quantities of protein were required before amino acid composition could be undertaken, and accuracy was poor.

Even today, when techniques such as ion-exchange chromatography are in use, the accuracy and reproducibility of analyses for cystine, methionine, and tryptophan are problematic, not least because of destruction of amino acids during the hydrolysis stage of analysis (see FAO/WHO 1991). Current views are that although the supply of any of the essential amino acids and of total nitrogen may theoretically limit protein utilization, only lysine, threonine, tryptophan, and the sulfur-containing amino acids are likely to be limiting in normal foods or diets. Even further, we now believe that in diets based on predominantly cereal protein sources, lysine may be the only amino acid that conditions overall protein quality (Pellett and Young 1988, 1990; Pellett 1996; Young, Scrimshaw, and Pellett 1997).

Digestibility of Protein

Although one would have expected the concept of digestibility to have developed early in the history of nutrition, a view of digestion and digestibility that we would recognize today did not arise until the middle of the nineteenth century. Following Hippocrates, the common view was that although there were many kinds of foods, there was but a single aliment. The belief of a single aliment prevailed throughout many centuries because little knowledge of the chemical nature of the various foodstuffs existed. Even in the early days of scientific nutrition research, William Beaumont (famous for his studies on digestion using a subject with a fistulous opening in his stomach caused by a gunshot wound), still held to this belief despite the then-developing opinions on the role of proteins, fats, and carbohydrates as chemically present in foods (McCollum 1957). In his own words, "the ultimate principles of aliment are always the same, from whatever food they may be obtained" (Beaumont 1833: 275) and "no other fluid produces the same

effect on food that gastric juice does; and it is the only solvent of aliment" (Beaumont 1833: 278).

Although the dictionary definition of digestibility – "capable of being digested" – embraces both the scientific and the popular conception of the term, scientifically the term should suggest a numerical proportion of the nutrient capable of being digested (absorbed). In contrast, popular usage of the word "digestibility" merely seems to imply that indigestion is being avoided. The concept of being absorbed applies to any nutrient, but in practice, the term digestibility is mainly employed for food energy and for protein, whereas "bioavailability" is used to describe the degree of absorption and/or utilization of other nutrients.

Digestibility considerations were recognized as important in studies published by Boussingault from the 1840s onward. He made a number of quantitative experiments to determine the rate at which different constituents of food disappeared from the digestive tract and at the same time tabulated the theoretical amounts of vegetable feeds that would produce equal effects on the growth of animal muscle. As an example, if a standard 10 pounds (lbs.) of hay were needed to produce a certain amount of muscle, only 2 lbs. of linseed oil cake and 5 lbs. of oats, respectively, were required to do the same. Or, again to produce the same growth, 52 lbs. of wheat straw and 61 lbs. of turnips were needed. These values were, in effect, comparative protein nutritional values of feeds and, as with the majority of our current biological techniques of protein evaluation, implicitly included digestibility as a component.

Boussingault (1850) was also the first to distinguish between nitrogen as ammonium salts and nitrogen as urea in urine and was instrumental in initiating nitrogen balance experiments in farm animals. His studies (see McCollum 1957) were followed at Rothamsted in England by the work of J. B. Lawes and J. H. Gilbert in 1859. They fed two pigs that were closely similar in weight, one with lentil meal containing 4 percent nitrogen and the other barley meal containing 2 percent nitrogen. The pig fed lentil meal excreted more than twice as much urea nitrogen as did the other fed barley meal. It was evident that the pigs on these two sources of protein retained very different percentages of their food protein for conversion into body protein.

Under the impetus of the naval blockades during World War I, Rubner, in Germany, took up investigations on the nutritive value of bread. His general conclusion, following a number of digestibility studies, was that flour milled at a rate above 70 percent was poorly used by humans. The nitrogen in bran was, however, an excellent food for cattle, and the high-fiber milling residue was therefore recommended for animal feed.

Although digestibility considerations were known to be important, and it was recognized that digestibil-

ity was inclusive in most biologically determined protein quality values, the concept suffered relative neglect for a long period, with amino acid composition data given far greater attention. Only recently has the relevance of digestibility as perhaps the more important component of protein quality (that is, compared to amino acid composition) again been emphasized (FAO/WHO 1991).

Amino Acid Metabolism

The development of the concepts of intermediary metabolism of protein and amino acids – that is, the stepwise conversion of these materials to the end products, urea, water, and CO_2 , with the liberation of energy – has been a continuously evolving process since the early years of the twentieth century. Hopkins in 1913 emphasized that “in the study of the intermediate processes of metabolism we have to deal, not with complex substances which elude ordinary chemical methods, but with simple substances undergoing comprehensible reactions” (Hopkins 1913: 653). Probably the major event in this progression of knowledge was the discovery of the Krebs-Henseleit cycle (Krebs and Henseleit 1932; Cohen 1981). The concept of the “cycle” became fundamental, not only for understanding the excretion of urea but also to the whole future of biochemistry. As Krebs remarked (see Estabrook and Srere 1981) at his 80th birthday symposium:

Perhaps the most novel and fundamental contribution was the discovery of the urea cycle. It preceded the tricarboxylic cycle for the Nobel Prize. It was very important to have the concept of “the cycle.” . . . And we know, of course, by now that there are many dozens, if

not hundreds, of cyclic processes (Estabrook and Srere 1981: xvii).

Dietary proteins are subjected in the body to a series of metabolic reactions, which are outlined in Figure IV.C.3.3. Following ingestion, protein foods undergo digestion, whereupon the amino acids are liberated by the digestive enzymes. Amino acids are then absorbed as free amino acids and as two and three amino acid compounds (di- and tripeptides). Nonabsorbed protein and/or nitrogen compounds appear in the fecal material but may have been subjected to further bacterial reactions in the large intestine. For most protein foods, nitrogen absorption is in excess of 90 percent. Following the distribution of the free amino acids between the extra- and intracellular amino acid pools, subsequent reactions can be considered in four categories:

1. Part of the free amino acid pool is incorporated into tissue proteins. Because of subsequent protein breakdown, these amino acids return to the free pool after a variable length of time and thus become available for reutilization for protein synthesis or for catabolism.
2. Part of the intracellular amino acid pool undergoes catabolic reactions, mainly in the liver. This process leads to loss of the carbon skeleton as CO_2 or its deposition as glycogen and fat, whereas the nitrogen is eliminated as urea via the Krebs-Henseleit cycle.
3. Nonessential amino acids are made in the body using amino groups derived from other amino acids (including the EAAs if they are in excess of requirement) and from carbon skeletons formed by reactions common to the intermediary metabolism of carbohydrate (Table IV.C.3.9).

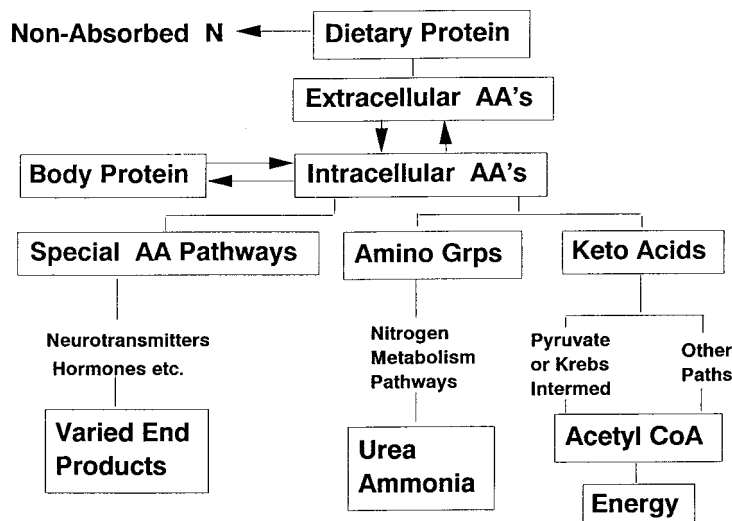


Figure IV.C.3.3. Metabolism of dietary protein. After the absorption of amino acids they are either incorporated into protein or are metabolized into α ketoacids and amino groups. Acetyl coenzyme A may either be used for energy or stored as fat. Pyruvate or Krebs cycle intermediaries may be used for the synthesis of glucose. Additional pathways involve the special role of amino acids for other uses as described in Table IV.C.3.10.

Table IV.C.3.9. *Summary of nonessential amino acid biosynthesis in mammals*

Amino acid	Precursor	Pathway
Alanine	Pyruvate	Transamination
Glutamate	α KG	Transamination
Glutamine (ATP/Mg)	α KG	GDH (+NH ₄)
Aspartate	Glutamate	+ NH ₄
Asparagine	OAA	Transamination
Serine	Aspartate	+ NH ₄
	3-P glycerate	Phosphoserine
Glycine	Serine	HO-pyruvate
Methionine	Homocysteine	THF needed
Cyst(e)ine	Serine + homocysteine	Betaine needed
Tyrosine	Phenylalanine	Cystathionine
Proline	Glutamate	Bioptherin need
Arginine	Glutamate	Semialdehyde
		Ornithine

Abbreviations: GDH glutamic dehydrogenase.

α KG alpha ketoglutaric acid

OAA oxaloacetic acid

THF tetrahydrofolate

Note: Methionine is an essential amino acid but can be synthesized from homocysteine, which, however, is not a normal dietary constituent.

4. Some free amino acids are used for synthesis of a number of new and important N-containing compounds. Examples are shown in Table IV.C.3.10 and include purine bases, creatine, gamma amino butyric acid, serotonin, and catecholamines. However, the portion of the total amino acid requirement devoted to this important function is small in absolute terms in comparison with the amount of amino acids needed for protein synthesis.

A summary listing of the end products formed from both the nitrogen and the carbons of the major amino acids as they are degraded is shown in Table IV.C.3.11, and a diagrammatic overview of the breakdown of the amino acids for energy and for other purposes is also illustrated in Figure IV.C.3.4.

Amino Acid Scoring Systems

Overlapping the development of animal assays for evaluating protein quality were significant advances in organic chemistry that allowed purification and isolation of both proteins and the amino acids. The next major advances in studies of protein metabolism were made by W. C. Rose and his associates at the University of Illinois. Their investigations, which began in the 1920s and extended until the 1950s, were able to distinguish, as unequivocally as is possible in the biological sciences, between the essential (indispensable) and the nonessential (dispensable) amino acids (Rose 1957).

Table IV.C.3.10. *Other functions of some amino acids*

Amino acid	Precursor of	Role
Alanine	Pyruvate	Amino transport Gluconeogenesis
Arginine	Creatine Ornithine Polyamines Oxaloacetic acid	Energy metabolism Urea formation Cell division? Gluconeogenesis
Aspartic acid	Aparagine	Ammonia transport
Cysteine	Purine, pyrimidines Glutathione	Nucleic acids Oxidation/reduction Amino acid transport
Glutamic acid	Taurocholic acid α -ketoglutarate	Fat absorption Gluconeogenesis, transamination
Glutamine	Glutamine Glutathione γ -amino butyric acid Purines, pyrimidines Amino sugars Citrulline	Ammonia transport Oxidation/reduction Neurotransmitter Nucleic acids Glycoproteins Ammonia transport
Histidine	Histamine	Biological amine
Lysine	Carnitine	Lipid transport
Methionine	S-adenosylmethionine	Methylation reactions
Phenylalanine	Tyrosine	Hormone synthesis
Serine	Ethanolamine, phosphoserine 4-sphingenine	Phospholipid synthesis Cerebroside synthesis
Tryptophan	Serotonin Niacin	Neurotransmitter Vitamin
Tyrosine	Melanin Catecholamines Thyroxin	Pigment Hormones

Note: These functions are additional to the roles as constituents of proteins and as energy sources.

The early studies were on rats fed rations containing mixtures of the then-known amino acids in place of protein. The rats did not grow, despite the fact that the concentrations of amino acids were similar to those in milk protein (casein). They did grow, however, when a small amount of intact casein was added. This suggested the presence of an unknown substance in casein, which was later identified as threonine, the last of the essential amino acids in proteins to be recognized.

By the omission of amino acids one at a time, the 10 essential amino acids required by the rat were identified, and their daily needs could be estimated. Similar techniques were subsequently extended to humans, with the help of graduate students who not only isolated or synthesized the amino acids but also served as experimental subjects for nitrogen balance studies.

As human amino acid requirements became known, comparisons were made with the amino acid composition of foods and diets. Subsequently, the capacity of various proteins to meet human protein and amino acid requirements was compared and evaluated, and the results were called protein or amino

Table IV.C.3.11. *Fate of the nitrogen and the carbon atoms in the degradation of the amino acids for energy*

Amino acid	Nitrogen path	Carbon path
Aspartate	Used for argine or appears as NH_4 in purine nucleotide cycle	Fumarate formed from both pathways. Oxaloacetate can also be formed through transamination
Glutamate and precursors		
Glutamate	As ammonia (GDH) or as aspartate (AAT)	As alpha-ketoglutarate
Glutamine	Amide N as NH_4	As glutamate
Proline	As Glutamate	As glutamate
Arginine	Two appear in urea, one in glutamate	One appears in urea and five C in glutamate
Ornithine (if excess is produced)	Terminal N transferred to α -KG to give glutamate	Glutamate semi-aldehyde to glutamate
Pyruvate and precursors		
Alanine	In glutamate by transamination	As pyruvate
Serine	Direct to NH_4 by SDH	As pyruvate
Cysteine	In glutamate by transamination	As pyruvate (sulfur as sulfate)
Threonine	Direct to NH_4 by SDH	Succinyl CoA (a homologue of pyruvate) via propionyl CoA
Branched chain amino acids		
Leucine	In glutamate by transamination	As acetoacetate via 3-hydroxy-3-methylglutaryl CoA
Isoleucine	In glutamate by transamination	CoA (via propionyl CoA) and acetyl CoA
Valine	In glutamate by transamination	As succinyl CoA via propionyl CoA
Aromatic amino acids		
Phenylalanine	In glutamate by transamination	As acetoacetate and fumarate
Tyrosine (formed from phenylalanine by hydroxylase)	In glutamate by transamination	As acetoacetate and fumarate
Tryptophan	One N appears as alanine, one as NH_4	Three C appear as alanine, four as crotonyl CoA, and one as formate and three as CO_2
Other amino acids		
Lysine	Both Ns transferred to glutamate	Two appear as CO_2 and four as crotonyl CoA
Glycine	As NH_4	As CO_2 and pyruvate. Originate from two glycine via one-carbon pathways
Histidine	One N and five C as glutamate, two N as NH_4	As glutamate and as methylidine- H_4 folate
Methionine	Released as NH_4 via homocysteine	Four carbons as succinyl CoA via cystathionine and propionyl CoA. Sulfur from homocysteine appears in cysteine

acid scores (Block and Mitchell 1946). The amino acid composition of egg protein was the first reference pattern proposed because it was known to be of high quality in animal-feeding studies. It did not, however, prove to be a suitable pattern, and many proteins scored poorly in relation to egg. This was because the levels of indispensable amino acids contained in egg were well above estimated human needs.

Direct comparison with human amino acid requirements soon followed as these data became available (FAO 1957; FAO/WHO 1973; FAO/WHO/UNU 1985; FAO/WHO 1991). The general expectation developed that not only should the score be able to predict the potential nutritional value of a food (or diet) for humans but that such a score should also correlate directly with the results of animal assays, such as net protein utilization (NPU). This did not always follow, and we now consider that the appropriate standard for dietary assessment for humans should be the

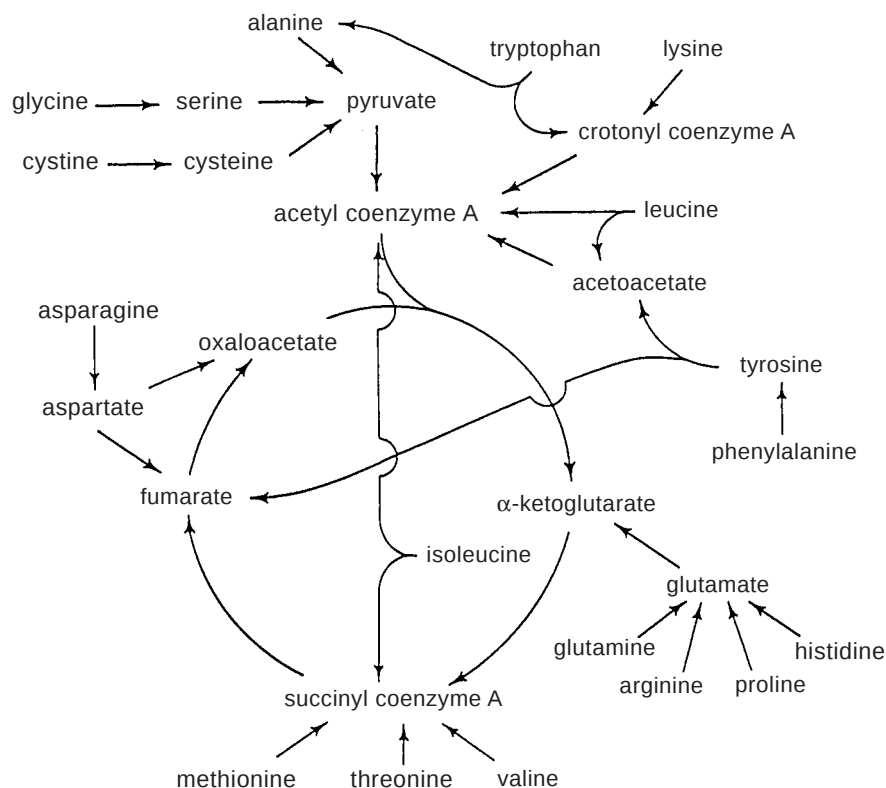
human amino acid requirement pattern. Some amino acid scoring patterns recommended for protein evaluation purposes in recent years are shown in Table IV.C.3.12.

Protein Requirements

The earliest estimates of protein requirements, made by L. Playfair (1853), Voit (1881), and Atwater (1891), were derived from observations of protein consumption, whereas the later studies, including those of R. H. Chittenden (1904), Sherman (1920) and E. F. Terroine (1936) were based on a more physiological approach in that estimates were made of the minimum amount of protein required to support body nitrogen equilibrium.

Actual protein requirement data have been debated for many years; even as early as 1919, Osborne noted that the amount of protein that should be included in

Figure IV.C.3.4. Metabolic pathways for the amino acids.



the human diet had long been a matter of debate among physiologists and nutrition experts, yet no agreement seemed to be in sight. Starting with a Health Committee of the League of Nations in 1936 and continuing to the present (IDECG 1996), estimates of human protein requirements and dietary protein allowances have been proposed by various national and international working groups. The major committees concerned with protein are shown in Table

IV.C.3.13. These estimates, or "standards," have been the basis for food and agricultural policies and programs, and for the planning and evaluation of diets for groups of individuals. The strengths and weaknesses of such international committee efforts to establish protein requirements have been examined by N. S. Scrimshaw (1976), who recognized that the recommendations reached by a committee are often less rigorous than those that would be proposed by an individual.

Table IV.C.3.12. Recommended scoring patterns, 1950-91 (mg/g protein)

	EGG	FAO (1957)	FAO/WHO (1973)	NAS-NRC (1974)	FAO/WHO/UNU (1985)			FAO/WHO (1991)		
					Infant	Child	School-age children	Adult	Infant	All other ages
Iso	54	42	40	42	46	28	28	13	46	28
Leu	86	48	70	70	93	66	44	19	93	66
Lys	70	42	54	51	66	58	44	16	66	58
Saa	57	42	35	26	42	25	22	17	42	25
Aaa	93	56	61	73	72	63	22	19	72	63
Thr	47	28	40	35	43	34	28	9	43	34
Trp	17	14	10	11	17	11	9	5	17	11
Val	66	42	50	48	55	35	25	13	55	35
TOTAL	490	314	360	356	434	320	222	111	434	320

Source: FAO/WHO (1991).

Table IV.C.3.13. *International protein recommendations (1936–96)*

1936	League of Nations: Health
1957	FAO: Protein
1965	FAO/WHO: Protein
1973	FAO/WHO: Protein/Energy
1985	FAO/WHO/UNU: Protein/Energy
1991	FAO/WHO: Protein Quality
1996	IDECG: Energy/Protein

Today, one might echo the sentiments Osborne penned in 1919, as each expert committee reports its findings. Nevertheless, agreement is now much closer, especially when we compare present recommendations (Table IV.C.3.14) with some of the high values for intakes and, hence, requirement recommendations proposed in the early 1900s (Table IV.C.3.6). These earlier data, especially those from the United States, are also remarkable for the very high levels of both fat and food energy recommended. In contrast, the recommendations from Chittenden (1907) are also shown in Table IV.C.3.6. Using himself as a subject, he consumed much lower levels of protein and maintained health. He did, however, eventually lose some weight. From his reported body weight his protein intake was 0.71 g/protein/kg/day, a little below the adult safe levels of 0.75 g/kg recommended by FAO/WHO/UNU (1985). In contrast, the Atwater (1891) recommendations for a man weighing 70 kg represented 1.6 g/kg/day.

Different committees have made different recommendations, sometimes based on the same original data (Irwin and Hegsted 1971). As an example, in a 1957 FAO report, the recommendations for adults were reduced significantly from the Sherman (1920) recommendations by applying different assumptions to the same data. These data extended back to the work of Atwater and others in the 1890s. At times, new recommendations have been generated based on data not available to earlier committees as, for example, with the young child data, originating from Guatemala, in a 1985 FAO/WHO/UNU report.

Revisions in the standards have had a profound effect on the interpretation of the possible protein nutritional status of a population. When, in the 1940s and 1950s, young child protein allowances were placed as high as 4.5 g/kg, a world protein problem was postulated. By revision of the protein requirement values downwards to about one-third of the previous allowances, total food energy rather than protein became the limiting factor, and the global “protein gap” disappeared at a stroke of a pen. These and other problems were discussed in scathing fashion by D. S. McLaren (1974) in an essay on “The Great Protein Fiasco.” A more recent example that influenced international agricultural policy was the recommendation from the FAO/WHO/UNU (1985) group concerning adult needs for essential amino acids. The implications from the recommendations (subsequently replaced) were that EAA needs could be fully met in the adult by a cereal protein such as wheat, and that the consumption of complementary proteins was no longer necessary.

The requirement for dietary protein consists of two major components: (1) total nitrogen to serve the needs for synthesis of the nutritionally dispensable (nonessential) and conditionally indispensable amino acids, as well as for other physiologically important nitrogen-containing compounds, and (2) the nutritionally indispensable (essential) amino acids that cannot be made by human tissues and so must be supplied from the diet. The most recent recommendations by a UN committee (FAO/WHO/UNU 1985) for all age groups are shown in Table IV.C.3.15. Comparative recommendations by various groups for selected ages over a number of years have already been shown in Table IV.C.3.14.

Protein requirements can be estimated in several ways (Pellett 1990b). One is to measure all losses of nitrogenous compounds from the body when the diet is devoid of protein and then assume that sufficient nitrogen from high-quality dietary protein to replace these obligatory nitrogen losses will provide the adult human subjects with their requirements. For the child, the needs for growth must also be considered. This is called the factorial method because it is based on

Table IV.C.3.14. *Summary of some recent committee recommendations for practical protein allowances in various age groups*

		Committee					
Age group	League of Nations 1936	FAO 1957	U.K. ¹ 1969	FAO/WHO 1973	NAS-NRC 1974	FAO/WHO/UNU 1985	IDECG 1996
g protein per kg per day							
Infants (1–2 years)	3.50	1.65	1.20	1.20	1.35	1.20	0.97
Teenage (15 years, boys)	1.50	0.69	0.57	0.72	0.67	0.87	0.84
Adults (35 years, men)	1.00	0.53	0.46	0.57	0.60	0.75	0.53

League of Nations (1936) allowances are expressed in terms of mixed dietary protein. All other allowances are in terms of high-quality, highly digestible protein.

¹United Kingdom Department of Health and Social Security

Table IV.C.3.15. *FAO/WHO/UNU (1985) safe levels of protein intake**

Age group	Males		Females
g protein/kg body weight [†]			
3-6 months	-	1.85	-
6-9 months	-	1.65	-
9-12 months	-	1.50	-
1-2 years	-	1.20	-
2-3 years	-	1.15	-
3-5 years	-	1.10	-
5-7 years	-	1.00	-
7-10 years	-	1.00	-
10-12 years	1.00	-	1.00
12-14 years	1.00	-	0.95
14-16 years	0.95	-	0.90
16-18 years	0.90	-	0.80
Adults	0.75	-	0.75
Pregnancy	-	-	+ 6 [‡]
Lactation 0-6 months	-	-	+ 17.5 [‡]
6 months +	-	-	+ 13 [‡]

*Assembled from FAO/WHO/UNU (1985) report.

[†]Uncorrected for biologic values (amino acid scores) of mixed dietary proteins for infants and children and for digestibility for all groups.

[‡]Additional protein per day (grams).

adding up a series of factors or fractions of total need, which represents maintenance requirements (obligatory nitrogen losses) as well as specific needs, such as those for growth in the child or milk production in the lactating female. The factorial approach as used by FAO/WHO (1973) is illustrated in Table IV.C.3.16. Enough protein to replace these losses and needs should meet requirements (Munro and Crim 1988).

A second procedure for estimating protein requirements is to determine directly the minimum amount of dietary protein needed to keep the subject in nitrogen equilibrium (balance). Ideally, the two methods should arrive at similar estimates of protein requirements. For infants and children, optimal growth, and not zero nitrogen balance, is the criterion used; similarly, the additional needs of pregnancy and lactation must be fulfilled. For the infant in the first months of life, however, human milk intake may be taken as equivalent to need. In an FAO/WHO/UNU (1985: 98) report, the basic assumption was that "protein needs of an infant will be met if the energy needs are met and the food contains protein in quantity and quality equivalent to breast milk."

The requirements for the nutritionally indispensable (essential) amino acids are currently a matter of considerable debate. This is largely because of the ¹³C-tracer studies reported by Young, D. M. Bier, and Pellett at MIT in 1989. The term "pattern of requirements" means the amount of a specific amino acid required per unit of protein need. Some amino acid requirement patterns have been shown in Table IV.C.3.12.

Such estimates have been criticized, in part because of the high food energy levels that were present in these diets (Young et al. 1989; FAO/WHO 1991) and have now been replaced by use of the FAO/WHO/UNU (1985) young child requirement pattern for all ages above infancy (FAO/WHO 1991).

Some average data on the amino acid composition of foods from the Massachusetts Nutrient Data Bank are shown in Table IV.C.3.17. Lysine is at a much lower concentration in cereals when compared with animal proteins. The sulfur-containing amino acids are distinctly lower in legumes and fruits, whereas threonine is lower in cereals and in fruits/vegetables. For the other indispensable amino acids, concentrations do not vary greatly, and these amino acids are also rarely in deficit in real diets. The nutritional significance of these data for humans is dependent on the choice of the reference amino acid requirement pattern used (Table IV.C.3.12). Clearly, if the FAO/WHO/UNU (1985) amino acid requirement pattern for the adult were to be used, it would indicate that no food protein, even from cereals, would show limitation and that complementation was no longer necessary to meet protein needs. Most cereal-based diets, however, would be lysine limited when compared to the FAO/WHO (1991) pattern. Because of these problems, use of the adult pattern from FAO/WHO/UNU (1985) has now been abandoned.

Table IV.C.3.16. *Factorial approach for human protein requirements: Adapted from FAO/WHO (1973)*

- Estimate obligatory N losses (O_N): from the sum of urinary, fecal, skin, and miscellaneous losses. These data are obtained from measurements made when subjects are consuming nonprotein diets.
- Estimate N requirements for growth (G_N), pregnancy (P_N), and lactation (L_N) from N content and amount of new tissue in growth, and in fetal and maternal tissues in pregnancy. For lactation, N content of milk is multiplied by average milk volume.
- Thus minimum nitrogen requirement for growth = $O_N + G_N$. Note: G_N will be zero for an adult and will vary with age for a child.
- Adjust for efficiency of N utilization: add 30% (not to be confused with protein quality; see G. below).
Requirement = $(O_N + G_N) \times 1.3$
- Adjust for individual variability (an additional 30%)
Safe Practical Allowance (SPA) + $[(O_N + G_N) \times 1.3] \times 1.3$
- Additional allowance for pregnancy (A_P) or lactation (A_L)
 $A_P = (P_N \times 1.3) \times 1.3$ or $A_L = (L_N \times 1.3) \times 1.3$
Adjustments made both for efficiency of nitrogen utilization (+30%) and for individual variability (+30%).
- Values for nitrogen as computed above are converted to protein ($N \times 6.25$). These values are in terms of proteins of high quality, such as egg or milk proteins, and further adjustments are necessary to increase allowances when average protein quality value is less than 100%.

$$\text{Thus: } \frac{\text{Recommended Protein Allowance}}{(\text{Adjusted for quality})} = \frac{\text{SPA}}{\text{Protein Quality}}$$

Table IV.C.3.17. *Amino acid composition of major food groups from the Massachusetts Nutrient Data Bank*

Amino acid	Food group				
	Animal Mean $\pm$ SD	Cereals Mean $\pm$ SD	Legumes Mean $\pm$ SD	Nut/Seeds Mean $\pm$ SD	Fruit/Veg Mean $\pm$ SD
N	1726	170	153	153	572
Iso	46.7 (4.7)	39.8 (4.6)	45.3 (4.2)	42.8 (6.1)	38.5 (10.8)
Leu	79.6 (6.0)	86.3 (26.3)	78.9 (4.2)	73.5 (9.0)	59.1 (19.6)
Lys	84.3 (7.1)	30.5 (9.8)	67.1 (3.8)	43.5 (12.7)	49.2 (13.3)
Saa	37.7 (3.3)	41.1 (4.3)	25.3 (2.8)	37.7 (11.7)	23.6 (7.2)
Aaa	74.9 (8.2)	83.0 (9.2)	84.9 (6.3)	88.0 (16.9)	64.0 (18.4)
Thr	43.4 (2.6)	33.6 (5.4)	40.0 (3.3)	37.9 (5.4)	35.1 (8.7)
Trp	11.4 (1.5)	12.1 (3.3)	12.3 (2.4)	15.4 (4.6)	10.8 (3.9)
Val	51.2 (5.6)	51.1 (6.9)	50.5 (4.0)	55.6 (10.3)	45.9 (12.6)

Note: SDs are calculated from the observations recorded in the data bank but not all are truly independent variables.

Major differences exist between the lysine supply in developing and developed countries (Pellett 1996), as is illustrated in Table IV.C.3.18. Because of these differences, Young, Scrimshaw, and Pellett (1997) have speculated that in some regions of the world, lysine fortification of wheat flour may offer potential for overcoming limitations in protein quality in the available dietaries. Lysine fortification of cereal-based animal feeds is common practice.

Protein-Energy Ratios

Human nitrogen metabolism is highly sensitive to altered dietary intakes of both protein and energy caused by substrate availability, as well as by hormone mediation, which can influence the status of body nitrogen and energy metabolism. Other areas of concern relate to protein-energy malnutrition (PEM). Severe deficiencies of protein and/or food energy, especially in association with infections, can lead to PEM. However, imbalance between protein and energy levels, as for example with excess food energy

together with low protein diets, can also produce metabolic abnormalities, such as the kwashiorkor form of PEM. The role of infection in the pathogenesis of PEM is mediated, at least in part, by hormonal changes, which can directly affect both protein and energy metabolism in previously well nourished individuals. The protein-energy ratio in formulas is of special importance in catch-up growth following infection (Dewey et al. 1996) and also for the needs of preterm infants (Tsang et al. 1993).

Body protein or nitrogen balance in the healthy mature human is maintained within relatively narrow limits, indicating a regulation that is achieved by varying the rate of N excretion in relation to changes in N intake and the individual's metabolic state. H. N. Munro (1951, 1964b), D. H. Calloway and H. Spector (1954) and Calloway (1975) have demonstrated that nitrogen balance is influenced by the intake of both protein and energy, even when that of energy is more than adequate; in consequence, estimates of protein requirements, based on nitrogen balance, can be affected by the level of energy available. Changes in

Table IV.C.3.18. *Mean values per capita for the availability of specific indispensable amino acids in developed and developing regions. Data for 1994*

Region	Amino acid							
	Per day (mg)				Per g protein (mg/g)			
	Lys	Saa	Try	Thr	Lys	Saa	Try	Thr
Developing ¹	2947 (841)	2160 (583)	693 (205)	2204 (509)	49 (7)	36 (2)	11 (1)	37 (2)
Developed and transitional ²	6149 (1172)	3619 (561)	1177 (195)	3799 (604)	64 (5)	38 (0.6)	12 (0.4)	40 (1)

¹Data for 61 countries.

²Data for 29 countries.

() = SD

Calculated from food balance sheet data (FAOSTAT 1996).

food energy intake, below or above energy needs, affect nitrogen balance. The improvement in nitrogen balance caused by an increase in energy intake, however, can be attenuated if protein is inadequate; conversely, the beneficial effects of an increase in protein can be inhibited by an inadequate energy intake. Published nitrogen balance studies relating the influences of protein on energy and energy on protein requirements have been reevaluated by Pellett and Young (1992), who postulate that the effects on N balance are continuous when either protein or energy are increased and dispute the proposals of Munro and Caloway that the relationships exist as a series of plateaus. In summary, therefore, nitrogen balance must be seen as a result of both protein and energy intakes, and each must be defined before use can be made of the data for establishing protein requirements.

Metabolic Regulation of Food Intake

The initiation and termination of feeding are complex processes that involve a large number of signals to the central nervous system. In humans, cultural and social conventions are significant modifiers of the impact of metabolic and physiological signals. Historically, the focus of research on the regulation of energy intake has been motivated by observation on the physiological constancy of adult body weight and concern with deviations that occur in obesity and starvation. Maintenance of energy balance appears to be a primary goal of the mechanism controlling feeding behavior (Anderson 1988). Because metabolic energy is ultimately derived from the three macronutrients – carbohydrates, fat, and protein – each has been investigated for its role in control mechanisms (Mayer 1955). The presence of breakdown products in the gastrointestinal tract may directly or indirectly provide signals to the brain. Nutrients and hormones passing the liver result in changes in plasma concentrations, which, in turn, may signal the brain either directly or indirectly through a complex system of neurotransmitters. M. I. Friedman (1995) has proposed that eating behavior is triggered by a signal that is tied to ATP concentrations in the liver and is carried to the brain via afferents in the vagus nerve. The brain organizes the information arising from the metabolism of food and directs feeding so that the intake of energy and macronutrients is quantitatively regulated.

Problems of Excess and Deficit

Classification of body size by use of the Body Mass Index (BMI), a version of weight-for-height (W/H^2 : weight in kg, height in meters), to assess for overweight and obesity has been in use since it was originally proposed by A. Quetelet in the latter part of the nineteenth century. More recently, it has been extended as a means of classifying undernutrition in adults (FAO 1994; James and Ralph 1994). Full agreement on the exact cut-off points to be used has not

Table IV.C.3.19. *A proposed classification using BMI (W/H^2)*

CED Grade III	<16.0
CED Grade II	16.0–16.9
CED Grade I	17.0–18.4
Normal	18.5–24.9
Overweight	25.0–29.9
Obese	30.0–40.0
Morbid obese	>40.0

Note: BMI (W/H^2) is defined using weight in kilograms and height in meters.

Source: Adapted from USDHHS (1988) and FAO (1994).

yet been reached, but a provisional classification that ranges from severe undernutrition to severe overnutrition is shown in Table IV.C.3.19.

Obesity

That obesity could be a life-threatening condition was recognized by Hippocrates, who observed in one of his aphorisms that “[p]ersons who are naturally very fat are apt to die earlier than those who are slender” (Lusk 1933: 8). Because of the significant public health implications of obesity (see Pi-Sunyer 1988 and USDHHS 1988), much research activity has centered on its causation and prevention, but disagreements remain on most major issues. Nevertheless, it is generally agreed that obesity must represent the end result of a cumulative positive error in energy balance and, hence, a failure in whatever mechanism is supposed to maintain the situation. The nature of the mechanism is not agreed upon, although many hypotheses have been advanced. On one extreme is the view presented by L. S. P. Davidson and R. Passmore (1969), who wrote that “since the immediate causes of obesity are overeating and under-exercising the remedies are available to all, but many patients require much help in using them” (Garrow 1974: 4). Conversely, the opposite end of the spectrum is typified by E. B. Astwood, who suggested that “obesity is an inherited disorder and due to a genetically determined defect in an enzyme; in other words that people who are fat are born fat and nothing much can be done about it” (Garrow 1974: 4).

Malnutrition and Wasting

Associations between wasting, illness, and food were made early in the history of medicine but moved from being solely descriptive to the experimental with the studies of Lavoisier in the late 1700s. Release of heat from organic fuels metabolized in the living body was demonstrated to follow the same pattern of combustion as was observed in vitro. Loss of tissue mass during starvation was shown to involve oxidation of endogenous fuel stores in order to generate the energy required for life in sustaining metabolic activities. By the mid-nineteenth century there was a general consensus that the diet must contain a source of energy in order to maintain health, which was expanded later to

the periodic, if not daily, need for some 40 essential nutrients. Health is characterized by a range of body compositions, tissue functions, and metabolic activities, and the fundamental aim of nutritional assessment is to gather information on these areas so as to compare with the healthy reference population (Heymsfield and Williams 1988). Such comparisons can allow for formal definitions of the undernourished.

Malnutrition and starvation are extreme conditions, and the terms were originally used in a rather synonymous sense. Lusk (1928) placed the study of starvation as fundamental to the study of nutrition and defined both terms:

Nutrition may be defined as the sum of the processes concerned in the growth, maintenance, and repair of the living body as a whole or of its constituent organs. An intelligent basis for the understanding of these processes is best acquired by a study of the organism when it is living at the expense of materials stored within itself, as it does in starvation. Starvation or hunger is the deprivation of an organism of any or all the elements necessary to its nutrition. Thus when carbohydrates and fats only are eaten, protein hunger ensues. Complete starvation occurs when all the required elements are inadequate. (Lusk 1928: 75)

We no longer use “starvation” for the absence of a single nutrient, and the word now usually refers to a chronic, inadequate intake of dietary energy. Clinical starvation is, however, typically accompanied by multiple nutrient deficiencies. Quite divorced from the realities of the developing world, most scientific studies of starvation have used normal-weight or otherwise healthy subjects. Such investigations permit the clearest description of the metabolic response to starvation. One particular study (Keys, Brozek, Henschel et al. 1950), in which volunteer subjects, starting at normal weight, received about half of their normal intake over a 6-month period, is an acknowledged classic of physiological research. L. J. Hoffer (1988) noted that modern studies of starvation at the cellular level may not be directly applicable to the physiological situation. Whole-body studies thus remain essential but cannot always be ethically performed. Therefore, experimental data, including observations on “professional fasters” reported from the earlier years of the twentieth century (e.g. Lusk 1928), continue to be relevant.

How long a human can survive without food depends largely on the size of the individual’s energy stores. Provided that water is supplied, there are many reliable records of healthy people who have survived 4 to 8 weeks with no food at all, and several of these were described by Lusk (1928). J. V. G. A. Durnin and Passmore (1967) have indicated that a healthy man or woman can fast completely for 2 weeks and suffer no permanent ill effects. Most well-nourished persons can afford to lose 25 percent of their body weight

without danger to life. This is about the level achieved by the volunteers in the Keys study (1950). Very few have survived a loss of 50 percent of their body weight, although a dog recovered completely following weight reduction from 26.3 to 9.7 kg (Lusk 1928).

The essential features of the metabolic response to starvation are altered rates and patterns of fuel utilization and protein metabolism aimed at minimizing fuel needs and limiting lean tissue loss. The nature of the starvation diet determines the pattern of hormone levels and fuel consumption. In total fasting (and in other states of carbohydrate deprivation), fat oxidation dominates in a metabolic setting of low insulin levels and ketosis. Fuel utilization by the brain, kidney, and muscle is markedly altered in total fasting. The brain switches from exclusively glucose to predominantly ketone body oxidation. The fuel of resting muscle switches from predominantly fatty acids to ketone bodies, finally returning, after weeks of total fasting, to fatty acid oxidation again (Hoffer 1988). Other changes involve a rapid loss of labile body protein, followed by a prolonged period during which further protein losses are minimized in association with a slowed rate of synthesis and breakdown of body protein.

Unlike experimental starvation, the clinical diseases of protein–energy malnutrition are greatly influenced by associated environmental and health variables. Such conditions are commonly accompanied by many dietary deficiencies in addition to protein or energy, and the evolution of the syndrome is critically influenced by the effects of physiological stress. Severe undernutrition, now termed chronic energy deficiency (CED), is of major consequence worldwide and can be assessed by the use of Body Mass Index (see Table IV.C.3.19). Among the consequences to populations so affected are more sickness, lower work capacity and, hence, lower incomes, reduced social activity, and an increased risk of low-birth-weight (LBW) babies (FAO 1994; James and Ralph 1994).

Nutritional marasmus and kwashiorkor are considered to be the extreme forms of protein–energy malnutrition (PEM). The causes of PEM are complex and involve factors other than food insufficiency (Latham 1990). It is still unclear why some children develop nutritional marasmus, whereas others manifest features of kwashiorkor, including edema (Waterlow 1984). There is no doubt, however, that the major cause of nutritional marasmus is inadequate food intake, especially energy to meet the requirements for both metabolism and normal growth (Latham 1990).

Kwashiorkor was originally described and explained by Cicely Williams (1933) as a nutritional disease of childhood associated with a maize diet. Subsequently, protein deficiency in the presence of relatively adequate energy intake was identified as the main cause (Waterlow 1955). This hypothesis was supported by animal experiments in which kwashiorkor-like features were produced using a protein deficient diet with excess food energy. However, following publication of “The Great Protein Fiasco”

(McLaren 1974), it became widely accepted that energy deficiency was a much greater public health problem than was protein deficiency.

More recently, aflatoxin poisoning (Hendrickse 1984) was proposed as the cause of kwashiorkor; few, however, would agree that it is the sole cause, but it may be a contributing factor. The etiology of the disease remains under debate, but most scientists and physicians consider that although other factors play a role, kwashiorkor is usually due to a deficiency of protein relative to food energy (calories), with the disproportion exacerbated by infection. Even if the dietary intake itself is balanced, infection and infestation may cause anorexia, reduce food intake and nutrient absorption and utilization, or result in nutrient losses (Latham 1990). Kwashiorkor is not confined to children. Previously well-nourished adults have been found to develop the disease (termed famine edema) during food shortages (Latham 1990) or when food is withheld during a period of catabolic stress.

Peter L. Pellett

Notes

1. Atwater was such an important figure in the history of nutrition in the United States that it would be remiss not to expand on his role as the essential founder of the present-day, widespread, and important United States Department of Agriculture (USDA) nutrition activities. A Centennial Celebration Symposium was held in Washington in 1993 and is reported in a special issue of the *Journal of Nutrition* (Vol. 124, No. 9S). The abstract of the Centenary Lecture by Kenneth Carpenter (1994b) on the life and times of W. O. Atwater is quoted directly:

Wilbur Atwater, the son of a Methodist minister, grew up in New England and was an undergraduate during the Civil War. He obtained his Ph.D. in agricultural chemistry at Yale in 1869, and spent the next two years in post-graduate study in Germany. In 1874 he was appointed to the Faculty of Wesleyan University, and was their first Professor of Chemistry. For two years (1875-77) he directed the first agricultural experiment station in the U.S. at Wesleyan. When it was transferred to Yale he gradually changed his interest to human nutrition, and how the poor could make more economical food choices, though he believed that they needed to maintain high-protein intakes if they were to remain productive. In 1887, with the passage of the Hatch Act, a second experiment station was established in Connecticut with Atwater as director. He also served for nearly 3 years as the first director of the Office of Experiment Stations in Washington, D.C. His next ambition was to have the first human calorimeter in the U.S. This was operating at Wesleyan by 1893, and his team was able to compare the net energy values of carbohydrates, fats and alcohol in sparing body tissues. He compiled data from many sources for the composition and digestibility of foods, and also organized the setting up of cooperative programs of nutritional studies, funded by the U.S. Department of Agriculture, in many States. His active life ended with a stroke in 1904 (Carpenter 1994b; 1707S)

Bibliography

- AMA (American Medical Association). 1951. *Handbook of nutrition*. Second edition. New York and Philadelphia, Pa.
- Anderson, G. H. 1988. Metabolic regulation of food intake. In *Modern nutrition in health and disease*. Seventh edition, ed. M. E. Shils and V. R. Young, 557-69. Philadelphia, Pa.
- Astwood, E. B. 1962. The heritage of corpulence. *Endocrinology* 71: 337-41.
- Atwater, W. O. 1891. American and European dietaries and dietary standards. *Storrs Agricultural Experimental Station 4th annual report*, 106-71. Middletown, Conn.
1899. Discussion of the terms digestibility, availability, and fuel value. *Storrs Agricultural Experimental Station 12th annual report*, 69-110. Middletown, Conn.
- Atwater, W. O., and F. G. Benedict. 1899. *Experiments on the metabolism of matter and energy in the human body*. Bulletin no. 69, U.S. Department of Agriculture. Washington, D.C.
- Atwater, W. O., and A. P. Bryant. 1900. Availability and fuel value of food materials. *Storrs Agricultural Experimental Station 12th annual report for 1899*, 73-110. Middletown, Conn.
- Barclay, A. 1937. *Pure chemistry: A brief outline of its history and development, Part 1 - historical review*. London.
- Beaumont, W. 1833. Experiments and observations on the gastric juice, and the physiology of digestion. Plattsburg, N.Y. Facsimile of the original edition reprinted on the occasion of the XIIIth International Physiological Congress, Boston, Massachusetts, 1929.
- Block, R. J., and D. Bolling. 1951. *The amino acid composition of proteins and foods*. Second edition, ed. Charles C. Thomas. Springfield, Ill.
- Block, R. J., and H. H. Mitchell. 1946. The correlation of the amino-acid composition of proteins with their nutritive value. *Nutrition Abstracts and Reviews* 16: 249-76.
- Calloway, D. H. 1975. Nitrogen balance of men with marginal intakes of protein and energy. *The Journal of Nutrition* 105: 914-23.
- Calloway D. H., and H. Spector. 1954. Nitrogen balance as related to caloric and protein intake in active young men. *American Journal of Clinical Nutrition* 2: 405-11.
- Carpenter, K. J. 1994a. *Protein and energy: A study of changing ideas in nutrition*. Cambridge and New York.
- 1994b. The life and times of W. O. Atwater (1844-1907). *The Journal of Nutrition* 124: 1707S-14S.
- Chittenden, R. H. 1904. *Physiological economy in nutrition*. New York.
1907. *The nutrition of man*. New York.
- Cohen, P. P. 1981. The ornithine-urea cycle: Biosynthesis and regulation of carbamyl phosphate synthetase I and ornithine transcarbamoylase. In *Biological cycles. Current topics in cellular regulation*, Vol. 18, ed. R. W. Estabrook and P. Srere, 1-19. New York and London.
- Conway, J. M., ed. 1995. Advances in human energy metabolism: Balancing energy requirements and energy intake. *American Journal of Clinical Nutrition*. 62 (5S): 1033S-179S.
- Davidson, L. S. P., and R. Passmore. 1969. *Human nutrition and dietetics*. Fourth edition. Baltimore, Md.
- Dewey, K. G., G. Beaton, C. Fjeld, et al. 1996. Protein requirements of infants and children. Supplement, ed. N. S. Scrimshaw, J. C. Waterlow, and B. Schürch. *Energy and Protein Requirements, European Journal of Clinical Nutrition* 50 (Supplement 1): S119-50.
- Durnin, J. V. G. A. 1996. Energy requirements: General principles. Supplement, ed. N. S. Scrimshaw, J. C. Waterlow,

- and B. Schürch. *Energy and Protein Requirements, European Journal of Clinical Nutrition* 50 (Supplement 1): S2-10.
- Durnin, J. V. G. A., and R. Passmore. 1967. *Energy, work and leisure*. London.
- Estabrook, R. W., and P. Srere, eds. 1981. *Biological cycles. Current topics in cellular regulation*, Vol. 18. New York and London.
- FAO (Food and Agricultural Organization of the United Nations). 1950. Calorie requirements. *FAO Nutrition Studies No. 5*. Washington, D.C.
1957. Protein requirements. *FAO Nutrition Studies No. 16*. Rome.
1994. Body mass index: A measure of chronic energy deficiency in adults. *FAO Food and Nutrition paper No. 56*, P. S. Shetty, W. P. T. James et al., Rome.
- FAO/WHO (Food and Agricultural Organization of the United Nations/World Health Organization). 1965. Protein requirements. Report of a joint FAO/WHO expert committee. *FAO Nutritional Studies No. 37*. Rome.
1973. Energy and protein requirements. Report of a joint FAO/WHO ad hoc expert committee. *WHO Technical Report Series No 522*. Geneva.
1991. Protein quality evaluation. Report of a joint FAO/WHO expert consultation. *FAO Food and Nutrition Paper No. 51*. Rome.
- FAO/WHO/UNU (Food and Agricultural Organization of the United Nations/World Health Organization/United Nations University). 1985. Energy and protein requirements. Report of a joint FAO/WHO/UNU expert consultation. *Technical Report Series No 724*. Geneva.
- FAOSTAT. 1996. Computerized information series. *Food Balance Sheets 1961-1994*, Food and Agriculture Organization. Rome.
- Fischer, E. 1901. Ueber die Hydrolyse des Caseins durch Salzsäure. *Zeitschrift für Physiologische Chemie* 33: 151-76.
- Friedman, M. I. 1995. Control of energy intake by energy metabolism. *American Journal of Clinical Nutrition* 62 (Supplement): 1096S-100S.
- Garrow, J. S. 1974. *Energy balance and obesity in man*. New York.
- Harris, J. A., and F. G. Benedict. 1919. *A biometric study of basal metabolism in man*. Carnegie Institute of Washington Publication 279. Washington, D.C.
- Hegsted, D. M. 1976. Energy needs and energy utilization. In *Present knowledge of nutrition*. Fourth edition, ed. D. M. Hegsted. Washington, D.C.
- Hendrickse, R. G. 1984. The influence of aflatoxins on child health in the tropics with particular reference to kwashiorkor. *Transactions of the Royal Society of Tropical Medicine and Hygiene* 78: 427-35.
- Heymsfield, S. B., P. C. Darby, L. S. Muhlheim, et al. 1995. The calorie: Myth, measurement and reality. *American Journal of Clinical Nutrition* 62 (Supplement): 1034S-41S.
- Heymsfield, S. B., and P. J. Williams. 1988. Nutritional assessment by clinical and biochemical methods. In *Modern nutrition in health and disease*. Seventh edition, ed. M. E. Shils and V. R. Young, 817-60. Philadelphia, Pa.
- Hill, J. O., C. Melby, S. L. Johnson, and J. C. Peters. 1995. Physical activity and energy requirements. *American Journal of Clinical Nutrition* 62 (Supplement): 1059S-66S.
- Hoffer, L. J. 1988. Starvation. In *Modern nutrition in health and disease*. Seventh edition, ed. M. D. Shils, and V. R. Young, 774-94. Philadelphia, Pa.
- Hutchison, R. 1903. *Food and the principles of dietetics*. New York.
- IDECG (International Dietary Energy Consultancy Group). 1996. International dietary energy consultancy group, Energy and protein requirements. Supplement, ed. N. S. Scrimshaw, J. C. Waterlow, and B. Schürch. *European Journal of Clinical Nutrition* 50 (Supplement 1).
- Irwin, M. I., and D. M. Hegsted. 1971. A conspectus of research on protein requirements of man. *Journal of Nutrition* 101: 385-430.
- James, W. P. T., and A. Ralph. 1994. The functional significance of low body mass index. *European Journal of Clinical Nutrition* 48 (Supplement 3): 51-202.
- Jones, D. B. 1941. Factors for converting percentages of nitrogen in foods and feeds into percentages of protein. *U.S. Department of Agriculture Circular No. 183*. Washington, D.C.
- Keys, A., J. Brozek, A. Henschel, et al. 1950. *The biology of human starvation*. Minneapolis, Minn.
- Kinney, J. M. 1988. Food as fuel: The development of concepts. In *Modern nutrition in health and disease*. Seventh edition, ed. M. E. Shils and V. R. Young, 516-32. Philadelphia, Pa.
- Kleiber, M. 1961. *The fire of life: An introduction to animal energetics*. New York.
- Krebs, H. A., and K. Henseleit. 1932. Untersuchungen über die Harnstoffbildung im Tierkörper. *Zeitschrift für Physiologische Chemie* 210: 33-66.
- Latham, M. C. 1990. Protein energy malnutrition. In *Present knowledge of nutrition*. Sixth edition, ed. M. L. Brown, International Life Sciences Institute, 39-46. Washington, D.C.
- League of Nations, Technical Commission of the Health Committee. 1936. *The problem of nutrition. Report on the physiological bases of nutrition*, Vol. 2. Geneva.
- Levine, R. 1978. Introduction: Modern development of the principles of nutrition. In *Diabetes, obesity and vascular disease: Metabolic and molecular interrelationships*, Part 1, ed. H. M. Katzen and R. J. Mahler, 1-5. New York and London.
- Lifson, N., G. B. Gordon, and McIntock. 1955. Measurement of total carbon dioxide production by means of D₂O¹⁸. *Journal of Applied Physiology* 7: 704-10.
- Lusk, G. 1914. *The fundamental basis of nutrition*. New Haven, Conn.
1928. *The elements of the science of nutrition*. Fourth edition. Philadelphia, Pa., and London.
1933. *Nutrition*. New York.
- Mayer, J. 1955. Regulation of energy intake and the body weight: The glucostatic theory and the lipostatic hypothesis. *Annals New York Academy of Sciences*. 63: 15-42.
- McCollum, E. V. 1957. *A history of nutrition*. Boston, Mass.
- McGilvery, R. W. 1979. *Biochemistry: A functional approach*. Philadelphia, Pa., and London.
- McLaren, D. S. 1974. The great protein fiasco. *The Lancet* 2: 93-6.
- Mendel, L. B. 1923. *Nutrition: The chemistry of life*. New Haven, Conn.
- Miles, C. W., P. Webb, and C. E. Bodwell. 1986. Metabolizable energy of human mixed diets. *Human Nutrition Applied Nutrition* 40A: 333-46.
- Mitchell, H. H. 1924. A method of determining the biological value of protein. *Journal of Biological Chemistry*. 58: 873-903.
- Moore, S., D. H. Spackman, and W. H. Stein. 1958. Chromatography of amino acids on sulfonated polystyrene resins. *Analytical Chemistry* 30: 1185-90.
- Moore, S., and W. H. Stein. 1954. Procedures for the chromatographic determination of amino acids on four percent cross-linked sulfonated polystyrene resins. *Journal of Biological Chemistry* 211: 893-907.

- Munro, H. N. 1951. Carbohydrate and fat as factors in protein utilization and metabolism. *Physiological Reviews* 31: 449-88.
- 1964a. Historical introduction: The origin and growth of our present concepts of protein metabolism. In *Mammalian protein metabolism*, Vol. 1., ed. H. N. Munro, 1-29. New York and London.
- 1964b. An introduction to nutritional aspects of protein metabolism. In *Mammalian protein metabolism*, Vol. 2, ed. H. N. Munro, 3-39. New York and London.
- Munro, H. N., and M. C. Crim. 1988. The proteins and amino acids. In *Modern nutrition in health and disease*. Seventh edition, ed. M. E. Shils and V. R. Young, 1-37. Philadelphia, Pa.
- National Academy of Sciences - National Research Council (NAS-NRC). 1974. *Improvement of protein nutriture: Food and nutrition board*. Washington, D.C.
- Needham, J., and E. Baldwin. 1949. *Hopkins and biochemistry, 1861-1947*. Cambridge.
- Osborne, T. B. 1919. What and how much should we eat. *Atlantic Monthly* 122: 332-41.
- Osborne, T. B., L. B. Mendel, and E. L. Ferry. 1919. A method of expressing numerically the growth promoting value of proteins. *Journal of Biological Chemistry* 37: 233-99.
- Pauly, P. J. 1990. The struggle for ignorance about alcohol: American physiologists, Wilbur Olin Atwater and the Woman's Christian Temperance Union. *Bulletin of Medicine* 64: 366-92.
- Pellett, P. L. 1990a. Food energy requirements in humans. *American Journal of Clinical Nutrition* 51: 711-22.
- 1990b. Protein requirements in humans. *American Journal of Clinical Nutrition* 51: 723-37.
1996. World essential amino acid supply with special attention to South-east Asia. *Food Nutrition Bulletin* 17: 204-34.
- Pellett, P. L., ed. 1963. Evaluation of protein quality. Publication 1100, National Academy of Sciences - National Research Council. Washington, D.C.
- Pellett, P. L., and V. R. Young. 1988. The contribution of livestock products to human dietary needs with special reference to West Asia and North Africa. Workshop on Small Ruminants in Mediterranean Areas, International Centre for Agricultural Research in Dry Areas (ICARDA), December 1987. Aleppo, Syria.
1990. Role of meat as a source of protein and essential amino acids in human protein nutrition. In *Meat and health*, ed. A. M. Pearson and C. E. Dutson, 329-70. London and New York.
1992. The effects of different levels of energy intake on protein metabolism and of different levels of protein intake on energy metabolism: A statistical evaluation from the published literature. In *Protein-energy interrelationships*, ed. N. S. Scrimshaw and Beat Schürch. International Dietary Energy Consultancy Group, Nestlé Foundation, 81-121. Lausanne, Switzerland.
- Pellett, P. L., and V. R. Young, eds. 1980. Nutritional evaluation of protein foods. *UNU/WHO Food and Nutrition Bulletin Supplement No. 4*, United Nations University. Tokyo.
- Pi-Sunyer, F. X. 1988. Obesity. In *Modern nutrition in health and disease*. Seventh edition, ed. M. E. Shils and V. R. Young, 795-816. Philadelphia, Pa.
- Prentice, A. M., C. J. K. Spaaij, G. R. Goldberg, et al. 1996. Energy requirements of pregnant and lactating women. Supplement, ed. N. S. Scrimshaw, J. C. Waterlow, and B. Schürch. *Energy and Protein Requirements, European Journal of Clinical Nutrition* 50 (Supplement 1): S82-111.
- Rose, W. C. 1957. The amino acid requirements of adult man. *Nutritional Abstracts and Reviews* 27: 631-47.
- Rothwell, N. J., and M. J. Stock. 1983. Diet-induced thermogenesis. In *Mammalian thermogenesis*, ed. L. Giardier and M. J. Stock, 50-63. London.
- Scrimshaw, N. S. 1976. Strengths and weaknesses of the committee approach: An analysis of past and present recommendations for protein in health and disease. *New England Journal of Medicine* 294: 136-42, 198-203.
- Seale, J. L. 1995. Energy expenditure measurements in relation to energy requirements. *American Journal of Clinical Nutrition* 62 (Supplement): 1042S-6S.
- Sherman, H. C. 1914. *Chemistry of food and nutrition*. New York.
1920. Protein requirement of maintenance in man and the nutritive efficiency of bread protein. *Journal of Biological Chemistry* 41: 97-110.
1952. *Chemistry of food and nutrition*. Eighth edition. New York.
- Shetty, P. S., C. J. K. Henry, A. E. Black, and A. M. Prentice. 1996. Energy requirements of adults: An update on basal metabolic rates (BMR's) and physical activity levels (PAL's). Supplement, ed. N. S. Scrimshaw, J. C. Waterlow, and B. Schürch. *Energy and Protein Requirements, European Journal of Clinical Nutrition* 50 (Supplement 1): S11-23.
- Spackman, D. H., W. H. Stein, and S. Moore, S. 1958. Automatic recording apparatus for use in the chromatography of amino acids. *Analytical Chemistry* 30: 1190-205.
- Terroine, E. F. 1936. The protein component of the human diet. *Quarterly Bulletin Health Organization, League of Nations* 5: 427-92. Geneva.
- Thomas, K. 1909. Über die biologische Wertigkeit der Stickstoffsubstanzen in verschiedenen Nahrungsmitteln. Beiträge zur Frage des physiologischen Stickstoffminimums. *Archiv für Anatomie und Physiologie, Physiologische Abteilung*, 219-302.
- Tsang, R. C., A. Lucas, R. Uauy, and S. Zlotkin. 1993. *Nutritional needs of the preterm infant*. Baltimore, Md., and London.
- USDHHS (U.S. Department of Health and Human Services). 1988. The surgeon general's report on nutrition and health. DHHS(PHS) Publication No. 88-50210. Washington, D.C.
- Van Slyke, D. D. 1911. The analysis of proteins by determination of the chemical groups characteristic of the different amino acids. *Journal of Biological Chemistry* 10: 15-55.
- Waterlow, J. C. 1955. Protein malnutrition. Proceedings of a conference in Jamaica 1953. Cambridge.
1984. Kwashiorkor revisited: The pathogenesis of oedema in kwashiorkor and its significance. *Transactions of the Royal Society of Tropical Medicine and Hygiene* 78: 436-41.
- Willcock, E. G., and F. G. Hopkins. 1906-7. The importance of individual amino acids in metabolism. *Journal of Physiology* 35: 88-102.
- Williams, C. D. 1933. A nutritional disease of childhood associated with a maize diet. *Archives Diseases of Children* 8: 423-33.
- Young, V. R., D. M. Bier, and P. L. Pellett. 1989. A theoretical basis for increasing current estimates of the amino acid requirements in adult man with experimental support. *American Journal of Clinical Nutrition* 50: 80-92.
- Young, V. R., N. S. Scrimshaw, and P. L. Pellett. 1998. Significance of dietary protein source in human nutrition: Animal and/or plant proteins. In *Feeding a world population of more than eight billion people: A challenge to science*, ed. J. C. Waterlow, 205-22. New York.



IV.D

Deficiency Diseases

IV.D.1 Beriberi

The complex of clinical disturbances long known as beriberi has been recognized since early in the twentieth century as arising because of a deficiency of thiamine. Like others in the group of B vitamins, thiamine has essential coenzyme functions in intermediary metabolism. The principal role of this water-soluble molecule is that of precursor of thiamine pyrophosphate (thiamine diphosphate), a coenzyme (often referred to as cocarboxylase) in energy generation through oxidative decarboxylation of alpha-ketoacids, such as pyruvic acid and alpha-ketoglutaric acid. Thiamine pyrophosphate serves also as the coenzyme for transketolase, a catalyst in the pentose phosphate pathway of glucose metabolism. Measurement of transketolase activity in erythrocytes and stimulation of activity by added thiamine pyrophosphate is the most convenient and sensitive method for detecting human thiamine deficiency. As the pyrophosphate and/or the triphosphate, thiamine appears also to have a role in facilitating conduction in peripheral nerves.

Like most vitamins, thiamine cannot be synthesized in the body and must be acquired in the diet. It is found in many foods and is most abundant in grains, legumes, nuts, and yeast. All meats and most dairy products contain some thiamine, but the richest sources are pork and eggs. Milk, however, is not a rich source. As a water-soluble vitamin, thiamine is easily lost in cooking water. Fish can supply good amounts, but with fermentation of raw fish, the enzyme thiaminase may alter the thiamine molecule, blocking its biological activity. Dietary thiamine is actively absorbed, mainly in the small intestine. The recommended daily dietary allowance is 1.2 to 1.5 milligrams for adult males and 1.0 to 1.1 milligrams for adult females, depending upon age, with a 50 percent increase in requirements during pregnancy and lactation. Up to about 25 milligrams can be stored by a healthy person, especially in heart muscle, brain, liver, kidney, and skeletal muscle.

With a deficient diet, skeletal muscle soon begins to lose thiamine, followed by the liver, nervous sys-

tem, and heart. Some intestinal bacteria are known to synthesize thiamine (Altschule 1978: 96), but this action probably contributes little, if at all, to human thiamine stores. Most thiamine excreted in urine is free, but some is also released as the pyrophosphate. Some thiamine may be lost in sweat, but the significance of this loss for workers in hot conditions remains to be clarified. Thiamine is low in toxicity, although very large doses may have some side effects. Large doses, however, are rapidly excreted, and so persistent toxic effects are rare.

The clinical consequences of advanced thiamine deficiency reflect, especially, the involvement of the cardiovascular system and the nervous system, both peripheral and central. The consequences of severe deficiency will certainly be disabling and may be fatal.

In the sections that follow, thiamine deficiency is first considered in the past, then in the present, with due attention to its distribution, to groups and populations at special risk, and to the range of clinical features. Cultural, socioeconomic, and behavioral factors, which influence risk, clinical responses, and outcomes, are noted. The final sections examine the history of research on etiology and of work leading to control and prevention.

Beriberi in the Past

Thiamine deficiency does not appear to leave behind any of the kinds of distinctive pathological evidence that permit paleopathologists to recognize diseases of the past. A comprehensive survey of pathological conditions in human skeletal remains (Ortner and Putschar 1985) includes no conditions assignable to deficiency of any of the B vitamins, and it seems unlikely that future studies of soft tissue remains will reveal any such evidence. Nevertheless, it is likely that thiamine deficiencies occurred in ancient populations, sporadically at least, even long before the beginnings of grain cultivation.

It has been shown that antithiamine factors occur in a variety of foods. Thermolabile thiaminases have been identified in the viscera of fresh fish, in shell-

fish, and in certain bacteria. Thermostable thiaminases exist in some animal tissues and in plants associated with caffeic and tannic acids (Paige 1988: 558). Some studies have shown an association of these factors with thiamine deficiency, and thus beriberi. S. L. Vimokesant and colleagues (1975), for example, have described thiamine deficiency in Thailand associated with the consumption of fermented raw fish, the chewing of fermented tea leaves, and the use of betel nuts. They suggest that the antithiamine activity of betel nuts and fermented tea may be induced by their high content of tannic acid. Ingestion of raw clams, which contain thiaminase, has also been shown to destroy thiamine in the human gastrointestinal tract. Some of these kinds of dietary preferences and practices have, no doubt, ancient origins in the human diet.

Early grain cultivation provided important new sources of dietary thiamine, and it is unlikely that simple milling procedures, even in rice-growing regions, led to severe thiamine deficiency, although diets based on rice stripped by steam milling and polishing of its thiamine-rich husks were later widely responsible for the disease. Ancient beriberi, as described in China, may more often have been due to thiaminase or other antithiamine activity than to the absence of the vitamin in the diet because of milling practices or dependence on thiamine-poor foods.

August Hirsch (1885: 572-3) records that beriberi, known as *kak-ke* ("disease of the legs"), can be traced back more than 2,000 years in China:

Inquiries . . . have succeeded in proving that the word "kak-ke" occurs in a Chinese work dating from about 200 B.C., and that there is an unambiguous description of the disease in another work of some one hundred and thirty years later date. Other references to beriberi in China occur in writings of the third, seventh, and eighth centuries of the present era; and in a medicine book belonging to the end of the tenth century, there is already a distinction drawn between a "dry" or paralytic kak-ke and a "wet" or dropsical [one]. Of the disease in Japan the earliest record is in a medical treatise of the ninth century of our era. . . . For the East Indies the first notice of it occurs in the medico-topographical treatise of Bontius [De Medicina Indorum by Jacob Bontius (1592-1631), a Dutch physician], who was acquainted with the disease there in the seventeenth century under its colloquial name "beriberi."

B. C. P. Jansen (1956: 2) provides a slightly earlier Dutch record: "Already in the year 1611 the first Governor-General of the Dutch East India Company wrote that in the Dutch East Indies there was a disease named beriberi which caused paralysis in the hands and feet."

Beriberi is commonly recorded in dictionaries, encyclopedias, and other modern sources as a term

derived from a word in Sinhala (Sinhalese), meaning weakness. Hugh M. Sinclair (1982), however, refers to Bontius's "excellent account of beriberi, the name meaning 'sheep,' since the partial paralysis of the disease caused patients to walk like sheep" - this, indeed, is the correct derivation of the word, as any comprehensive Malay-Indonesian dictionary will show. In *Kamus Pelajar Federal* (1973), for example, the Malay word *biri-biri* is assigned two meanings: (1) sheep, and (2) beri-beri (the disease). Hirsch (1885: 569), writing as early as the 1860s, casts doubt on alternative suggestions that the word beriberi derived from "Cingalese" or "Hindustani" words. Hirsch leaned toward a Malay origin, citing Bontius (who offered a mistaken transliteration of *biri-biri* - "bharyee" - a word said to mean "sheep"). Hirsch also noted another early Dutch author's suggestion that a Malay language word, "biribi" (obviously *biri-biri*, pronounced rapidly), meaning an "abrupt and tripping gait," was the origin of the term beriberi.

Thiamine deficiency remained an endemic problem until the second half of the nineteenth century. Although present in many parts of the world, beriberi usually appeared as a sporadic disease in individuals, groups, or populations at special risk because of "restricted," thiamine-poor diets. However, with the introduction of the steam-powered rice mill (about 1870), the disease became widespread and at times epidemic in some populations, especially in eastern and southeastern Asia. As a product of industrial and technological expansion, rice - commercially milled and relatively inexpensive - became generally available in a time of maritime competition and increasing establishment, in many societies, of prisons, asylums, military installations, and other institutions that offered only restricted diets.

In the closing decades of the nineteenth century in Japan, for example, beriberi (usually known as *kakke*) was present throughout the country, but was especially prevalent in military garrisons and on warships and naval training ships (Williams 1961). In Malaya, too, at the turn of the century, beriberi was widely distributed but especially conspicuous in closed institutions, such as jails and asylums for the mentally ill (Byron 1951).

It may be noted, however, that most populations in the Indian subcontinent, and Indian immigrant groups generally, remained almost free of beriberi during this period of technological, military, and institutional expansion - as they still do today. This is understandable because, throughout this region, rice has long been parboiled before milling. The unhusked rice, after preliminary soaking, is boiled in domestic vessels or steamed under pressure in the larger mills. The B vitamins are fixed in the grain so that they are not removed with the bran, and remain behind in the milled rice. The parboiling process has long been preferred because the woody outer husk is split, rendering its removal easier during milling (Davidson and Passmore 1966: 261). Beriberi has been endemic in

only one Indian state, along the coast of Andhra Pradesh north from above Madras to Vishakhapatnam, and this, according to Sir Stanley Davidson and R. Passmore (1966: 412), is the only area in India where raw-milled rice is the principal cereal.

The vast literature on beriberi reflects its profound impact on morbidity and mortality during the late nineteenth and early twentieth centuries. But this literature also reveals the confusion that existed about the disease, its identity, its causation, and its clinical effects. Beriberi had acquired many other names – regional, local, and culturally specific – in previous centuries, but cosmopolitan medical practitioners and researchers added a considerable variety of their own terms (for example, endemic multiple neuritis, pan-neuritis epidemica, polyneuritis epidemica, ship beriberi, infantile beriberi). These reflected the notion that beriberi might be several quite different diseases, an idea created by the diverse clinical effects of thiamine deficiency and by the common occurrence of several or multiple nutritional deficiencies in the same thiamine-deficient patient.

Beriberi Today

Thiamine deficiency disease persists in the modern world, but its epidemiology is very different from that of a century ago. Epidemic classical beriberi is rarely seen today. Although signs and symptoms of famine-related thiamine deficiency occur all too often, they are likely to be accompanied by clinical evidence of multiple deficiencies of vitamins and other nutrients. War, disaster, and drought-induced famines will continue into the future as contributors to deficiencies of thiamine, but beriberi as a distinct clinical entity will probably seldom occur in those circumstances.

As an endemic disease, however, beriberi has not disappeared. Some beriberi still occurs where rice is the staple cereal, polished rice is preferred, parboiling is not practiced, and few other sources of thiamine exist in the diet. In other areas, where rice is not the staple, beriberi may occur occasionally when refined carbohydrate foods are dominant in the diet. Davidson and Passmore (1966: 412), for example, mention the common occurrence of beriberi early in this century in the fishing communities of Newfoundland and Labrador. Often cut off during the long winter from outside provisions, people in these areas depended upon winter stores consisting mainly of flour made from highly milled wheat. Beriberi probably continues to appear today in similar circumstances of prolonged and extreme isolation.

In some countries, beriberi is so intimately associated with culturally maintained dietary preferences and practices that it is likely to persist well into the future. In northern Thailand, as noted, consumption of fermented raw fish has been reported to be an important cause of beriberi. Vimokesant and colleagues (1975) also concluded that antithiamine activity of fer-

mented tea and betel nuts must contribute to persistent thiamine deficiency in that region. However, to what extent these effects may result from destruction of thiamine, or from loose binding and partial blocking of activity, remains unclear. Further evaluation of these findings is warranted because of the wide use of these substances not only in Thailand but elsewhere in Asia. If the extensive use of betel nut in some regions of India interferes with thiamine activity or retention, it appears, nevertheless, that the thiamine-preserving effect of rice parboiling can maintain a sufficient thiamine balance to minimize the appearance of signs and symptoms of deficiency. This too is, however, a topic in need of further study, particularly in relation to the occurrence and significance of subclinical or borderline thiamine deficiency.

In addition to the problem of endemic beriberi, thiamine deficiency occurs sporadically in many countries, especially as a result of chronic alcoholism, but also in association with some chronic illnesses and with long-term parenteral nutrition (Altschule 1978: 98). As Richard S. Rivlin (1988) reports, thiamine absorption is exquisitely sensitive to ingested ethanol, which can significantly interfere with absorption. Alcoholics may fail to absorb most of their dietary thiamine – thiamine which is likely to be limited, in any case, as the diet is increasingly impaired and restricted. Rivlin notes that about 25 percent of alcoholics admitted to general hospitals in the United States have evidence of thiamine deficiency by either clinical or biochemical criteria. He cites evidence as well that alcohol can adversely affect the intermediary metabolism of thiamine, and that chronic liver disease, resulting from alcoholism, may impair the conversion of thiamine to thiamine pyrophosphate.

Clinical Features

The clinical distinction between wet and dry forms of beriberi can be traced back at least to the ninth century in Japan (Henschen 1966: 170) and to the tenth century in China (Hirsch 1885). The wet-dry differentiation has continued to appear in clinical classifications of thiamine deficiency disorders to the present day, with infantile beriberi accorded separate status in most breakdowns. Some have chosen to describe the disease in terms of three clinical clusters characterized by edema, nervous system signs and symptoms, and symptoms of disorder in the cardiovascular system (Meade 1993). Sinclair (1982) offers the classification adopted here. He sees severe thiamine deficiency giving rise to the following forms of beriberi:

1. Acute beriberi – including acute cardiac beriberi and infantile beriberi, along with Wernicke's encephalopathy (as the acute or subacute component of the Wernicke-Korsakoff syndrome).
2. Chronic beriberi – including dry beriberi, wet beriberi, and chronic beriberi heart (and here

Korsakoff's psychosis may be added as the other, more chronic component of the Wernicke-Korsakoff syndrome).

Acute Cardiac Beriberi

The picture of acute cardiac beriberi is that of congestive failure and cardiac overactivity manifested by dyspnea, orthopnea, precordial pain, palpitations, tachycardia, and edema. In its most fulminant form it has been known in East Asia as *Shoshin* beriberi. These acute forms of cardiac beriberi, uncommon today, were well known to clinicians in Asia a century ago; many patients died of heart failure. With intravenous thiamine, acute cardiac beriberi can readily be treated, with dramatic results.

Infantile beriberi, a condition in infants who are breast-fed by thiamine-deficient mothers, is, as Sinclair (1982) notes, very similar to acute cardiac beriberi in adults. Once common in eastern Asia, the condition usually occurs very early in life, with sudden onset and rapid progression to death. The anorexic, restless infant shows signs of developing cardiac failure, including tachycardia, dyspnea, edema, and aphonia, responsible for a characteristic grunt (the beriberi cry). Although less often recognized today, infantile beriberi, like cardiac beriberi in the adult, will respond well to intravenous or intramuscular thiamine. Improvement of the nursing mother's thiamine nutrition is also essential.

Wernicke's Encephalopathy

First described by Carl Wernicke in 1881, Wernicke's encephalopathy is an acute or subacute cerebral form of beriberi, only recognized as a consequence of thiamine deficiency during and just after World War II, in prisoner-of-war camps. Today it is a disorder associated principally with chronic alcoholism. The pathological changes are characteristically localized, involving the optic nerve, the gray matter around the third and fourth ventricles, and the aqueduct of Sylvius. Clinical features include nystagmus, diplopia, ophthalmoplegia, anorexia, vomiting, insomnia, mental apathy and disorientation, hallucinations, and variable loss of memory. Ataxia appears occasionally. In many cases, signs of peripheral neuropathy are also evident. The condition responds well to treatment with thiamine injections.

Dry Beriberi

In this once common clinical variety of beriberi, it is primarily the peripheral nervous system that is disordered. Dry beriberi is an ascending symmetrical bilateral peripheral neuritis. Early in the course, the complaints may be vague, beginning with stiffness, leg cramps, and numbness in the feet and ankles. Knee and ankle reflexes are first increased, then diminished, and finally lost as the disease advances. Muscle weakness spreads upward in the lower extremities. Affected muscles become tender, painful, numb, and

eventually atrophic, sometimes with hyperextension of the knee joints. Ataxia follows, with a characteristic gait (which led to the coining of the name "beriberi," meaning sheep). Hands and arms may also be affected, but not usually until signs and symptoms in the legs are pronounced. Thiamine therapy will arrest the progression and relieve the patient of most symptoms, but all of the damage in advanced disease may not be fully reversible.

Wet Beriberi

This chronic form of beriberi is primarily an edematous condition accompanied, in varying degree, by signs and symptoms of cardiac malfunction. Renal function generally remains intact. Early in the course, edema usually begins in the feet and legs, then extends upwards to the abdomen (producing ascites), thorax, pericardium, and face. Muscle wasting, if present, may be masked by edema. With pulmonary congestion, the chambers of the right heart dilate and the heart begins to fail. Anorexia is invariable, vomiting and diarrhea may occur, and thus the body begins to starve. The patient with wet beriberi is always at risk of a sudden increase in edema, circulatory failure, severe dyspnea, and death.

Chronic Beriberi Heart

Cardiac damage can occur as a result of thiamine deficiency without other signs and symptoms of beriberi. This chronic condition, now referred to as "beriberi heart," is most often associated with chronic alcoholism and/or a history of marked nutritional deficiency. When these causal factors are present, the diagnosis can be suspected when unexplained tachycardia, palpitations, venous congestion, characteristic electrocardiographic changes, and right heart enlargement are noted. Edema may also appear, but not with the same prominence as in cases of wet beriberi.

Causal Research

As beriberi extended its geographical range, incidence, and prevalence in the second half of the nineteenth century, it attracted increasing attention from medical investigators, who began the long process of causal research that extended into the first decades of the twentieth century. Hirsch, in his great *Handbook of Geographical and Historical Pathology*, first published between 1860 and 1864, gave many pages to a description of the disease and to a review and analysis of etiologic possibilities. He found it necessary to discard as principal causes climate, soil, lifestyle, and occupation, among other factors. He could only conclude that beriberi must arise as a result of some intoxication, that is, exposure to a poison. Later the idea emerged that beriberi might appear because of infection, and, until late in the century, theorizing was directed principally to these two causal possibilities.

The search for an infectious agent or a chemical poison, probably in food, continued throughout the remainder of the nineteenth century, but no one interested in beriberi and diet in this period considered the possibility of a deficiency of some specific substance (Williams 1961: 16). Two naval officers (the Dutch naval doctor, F. S. van Leent, and the chief of the Japanese naval medical service, K. Takaki) were successful in reducing the incidence of beriberi on shipboard by broadening diets previously limited largely to rice (Jansen 1956; Williams 1961). These achievements, pointing clearly to some dietary factor in beriberi, were not widely known or acknowledged at the time.

The first step toward the recognition of the role of thiamine deficiency in beriberi came in 1890 with the publication of a report by Christiaan Eijkman, a Dutch military surgeon in Java, of his studies of polyneuritis in domestic fowl. His experiments were initiated to uncover a presumed infectious cause of the disease by inoculating laboratory chickens with supposedly infectious material from beriberi patients. These birds, fed on crude rice, remained healthy, but when they were later switched to a diet of boiled and polished white rice, left over from the hospital, they began to show signs of polyneuritis not unlike that in human beriberi.

Still later, the chickens were returned to a crude rice diet, after the hospital chief forbade the use of the polished luxury rice for laboratory animals. The neuritis soon disappeared. Intrigued by this change, Eijkman fed two groups of fowl, one with crude rice, the other with polished rice. The latter group developed the same beriberi-like neuritis. The result of this simple experiment was decisive, not only in the history of beriberi but also as the beginning of the much wider field of research on vitamins and specific nutritional deficiencies. Eijkman expanded his investigations in Java until 1896, when he returned to Holland, eventually to a professorial appointment and to the award in 1929 (one year before his death) of a Nobel Prize in Physiology and Medicine, “for his discovery of the antineuritic vitamin.”

Eijkman's experimental work with fowl in Java was continued by another Dutch scientist, Gerrit Grijns, who soon became the first to express the theory “that in food there occur in small quantities unknown organic substances essential for health” (Jansen 1956). Grijns called them protective substances. He was able to extract such a substance, soluble in water, from rice bran and to employ it in the treatment of human beriberi. Later Casimir Funk coined the name “vitamine” (later shortened to “vitamin”) for a protective substance, an amine that he had isolated and assumed (wrongly as it turned out) to be the same as that studied by Eijkman and Grijns (Byron 1951; Jansen 1956).

The studies in Java were confirmed and extended in Malaya early in the new century (Byron 1951). W. L.

Braddon, who published much evidence to show that the cause of beriberi was connected intimately with the consumption of polished white rice, could only explain his findings as due to a poison in the rice. His work provided a foundation for three scientists, H. Fraser, A. T. Stanton, and W. Fletcher, at the Institute for Medical Research in Kuala Lumpur. During the decade following 1905, in their studies of road workers (and in Fletcher's study of patients at an asylum), these investigators further demonstrated the significance of rice polishing in the causation of beriberi. Fraser and Stanton also failed in all attempts to isolate a poison from rice, and they concluded, as Grijns had earlier, that the answer to the problem of beriberi must lie in a nutritional defect, that is, in the absence of an essential protective substance. Fraser and Stanton were never able to isolate the substance, but their studies led them to call for, and advocate, practical measures to cure and prevent beriberi – measures that were to prove very successful. After 1911, the use of polished white rice was prohibited in all Malayan government institutions, the danger of polished rice consumption was widely publicized, and government rice mills began to produce undermilled rice for use in hospitals and public institutions, including prisons and asylums. These actions served to lower dramatically the incidence of beriberi in Malaya within just a few years.

After 1910, as the results of the Malayan studies and preventive actions became known, E. B. Vedder and others in the Philippines started similar programs for the prevention and treatment of the disease. Vedder also encouraged Robert Williams, a scientist at the Bureau of Science in Manila, to begin studies that were to lead, in 1933, to the isolation and final chemical characterization of the protective substance. Some years earlier, in Java, Jansen and W. F. Donath had succeeded in isolating crystals of the antiberiberi factor, but their chemical analysis, limited by the apparatus then at hand, was incomplete, missing the presence of sulfur in the molecule (Jansen 1956). After the final synthesis, in 1936, Williams named the chemical “thiamin.” The spelling was later changed to thiamine, but Williams's original spelling is still often used.

Prevention

Southeast and East Asian countries pioneered in developing strategies for the prevention of beriberi. As we have seen, prevention of shipboard beriberi through dietary change was initiated by Japanese and Dutch naval medical officers late in the nineteenth century, long before the recognition of thiamine's role. The section “Causal Research” also took note of the development, early in the twentieth century, of countrywide preventive programs and policies, begun in Malaya and the Philippines many years before thiamine was isolated, characterized, and synthesized.

These programs stressed changes in rice milling to preserve bran, together with public education about the risks associated with dependence upon a diet based on polished white rice.

After 1936 the pace quickened as production of inexpensive synthetic thiamine made food enrichment possible. By 1941 the first standards for the enrichment of bread were established in the United States, and enrichment requirements soon followed, during World War II. Enrichment of whole-grain rice presented more problems than flour enrichment for bread, but even these difficulties were overcome so that today thiamine enrichment is available in rice-based diets, as well as in those for which bread is a staple.

For people throughout the world who consume many other kinds of processed foods, including corn and macaroni products, thiamine is also commonly available, as it is in many varieties of vitamin preparations. In the United States today, many individuals are consuming vitamins, often including thiamine, far in excess of the recommended dietary allowances, and it is estimated that one-third of individuals 65 years of age and older in this country are taking some kind of nutritional supplement (Rivlin 1988).

Williams concluded his 1961 book, *Toward the Conquest of Beriberi*, with a chapter on the prospects for eradication of the disease, as advocated by a joint United Nations committee (World Health Organization 1958). This committee, meeting under the sponsorship of the Food and Agriculture Organization and the World Health Organization, recommended six preventive measures that, taken together, could be expected largely to eliminate, if not actually eradicate, beriberi in countries where rice is a staple food:

1. General improvement of the diet to increase its thiamine content.
2. Encouragement of the use of undermilled rice.
3. Encouragement of the use of parboiled rice.
4. Promotion of methods of preparing and cooking rice to preserve its thiamine content to the greatest extent possible.
5. Enrichment of rice.
6. Increased use of synthetic thiamine in various ways other than rice enrichment.

Each of these measures continues today to have its role in beriberi prevention. It should be noted, however, that the committee was not aware, in the 1950s, of the possible extent of the problem of interference with thiamine activity and availability through the actions of thiaminase and other antithiamine agents. Even now this problem remains largely undefined, although in some countries, certainly in northern Thailand and, probably, in areas of Myanmar (Burma), dietary preferences, especially for fermented raw fish, are considered to be the most important causes of endemic beriberi.

Sporadic and subclinical or marginal thiamine deficiency persists today and will continue to do so. Eradication of this deficiency is not in prospect, but neither is recurrent epidemic beriberi.

Frederick L. Dunn

Bibliography

- Altschule, Mark D. 1978. *Nutritional factors in general medicine. Effects of stress and distorted diets*. Springfield, Ill.
- Byron, F. E. 1951. Beri-beri. In *Fifty years of medical research in Malaya*. Studies from the Institute for Medical Research, Federation of Malaya, No. 25: 98-118. Kuala Lumpur.
- Davidson, Stanley, Sir, and R. Passmore. 1966. *Human nutrition and dietetics*. Third edition. Edinburgh and London.
- Henschen, Folke. 1966. *The history of diseases*, trans. Joan Tate. London.
- Hirsch, August. 1885. *Handbook of geographical and historical pathology*, Vol. 2, trans. Charles Creighton. London.
- Jansen, B. C. P. 1956. Early nutritional researches on beriberi leading to the discovery of vitamin B1. *Nutrition Abstracts and Reviews* 26: 1-14.
- Kamus Pelajar Federal. 1973. *Learner's Dictionary*. Kuala Lumpur and Singapore.
- Meade, Melinda S. 1993. Beriberi. In *The Cambridge world history of human disease*, ed. K. F. Kiple, 606-12. Cambridge and New York.
- Ortner, Donald J., and Walter G. J. Putschar. 1985. *Identification of pathological conditions in human skeletal remains*. Smithsonian Contributions to Anthropology No. 28. Washington, D.C.
- Paige, David M. 1988. *Clinical nutrition*. Second edition. St. Louis, Mo.
- Rivlin, Richard S. 1988. Vitamin B1 (thiamin) structure and biochemical function. In *Cecil textbook of medicine*. Eighteenth edition, 1229-30. Philadelphia, Pa.
- Sinclair, Hugh M. 1982. Thiamin. In *Vitamins in medicine*, Vol. 2. Fourth edition, ed. Brian M. Barker and David A. Bender, 114-67. London.
- Vimokesant, S. L., D. M. Hilker, S. Nakornchai, et al. 1975. Effects of betel nut and fermented fish on the thiamin status of northeastern Thais. *American Journal of Clinical Nutrition* 28: 1458-63.
- Williams, Robert R. 1961. *Toward the conquest of beriberi*. Cambridge, Mass.
- World Health Organization. 1958. *Fifth report of the Joint FAO/WHO Expert Committee on Nutrition*. Geneva.

IV.D.2 Iron Deficiency and Anemia of Chronic Disease

Until the nineteenth century, unspecified chronic anemia was known as chlorosis, or the "green sickness," referring to the extreme pallor that characterized severe cases. For centuries, "chlorosis, or green sick-

ness, was attributed to unrequited passion. Medieval Dutch painters portrayed the pale olive complexion of chlorosis in portraits of young women" (Farley and Foland 1990: 89). Although such extreme cases are not common in Western societies today, less severe acquired anemia is quite common. In fact, acquired anemia is one of the most prevalent health conditions in modern populations.

Technically, anemia is defined as a subnormal number of red blood cells per cubic millimeter (cu mm), subnormal amount of hemoglobin in 100 milliliter (ml) of blood, or subnormal volume of packed red blood cells per 100 ml of blood, although other indices are usually also used. Rather than imputing anemia to unrequited love, modern medicine generally imputes it to poor diets that fail to replenish iron loss resulting from rapid growth during childhood, from menstruation, from pregnancy, from injury, or from hemolysis. One of today's solutions to the frequency of acquired anemia is to increase dietary intake of iron. This is accomplished by indiscriminate and massive iron fortification of many cereal products, as well as the use of prescription and nonprescription iron supplements, often incorporated in vitamin pills. However, a nutritional etiology of anemia as dietary has, in the past, been assumed more often than proven. Determining such an etiology is complicated by the fact that the hematological presentation of dietary-induced iron deficiency anemia resembles the anemia of chronic disease. Of the many types and causes of acquired anemia, only those associated with diet and chronic disease are discussed here (for an overview of others, see Kent and Stuart-Macadam this volume).

Anemia Due to Diet or Chronic Disease?

Although causes differ, patients with either iron-deficiency anemia or the anemia of chronic disease/inflammation have subnormal circulating iron levels called hypoferremia. Below-normal levels of circulating iron are manifested in low hemoglobin/hematocrit, serum iron, and transferrin saturation levels. Because serum ferritin is an indirect measure of iron stores, it provides a sensitive index to distinguish these two anemias (Cook and Skikne 1982; Zanella et al. 1989). When the body does not have enough iron as a result of diet, bleeding, or other causes, serum ferritin values are subnormal and reflect the subnormal amount of iron in the bone marrow. In the anemia of chronic disease/inflammation, however, serum ferritin levels are normal to elevated because circulating iron is transferred to storage, as reflected by the combination of subnormal circulating iron levels with normal or elevated serum ferritin values.

Removing iron from circulation reduces its availability to pathogens that require it for proliferation (Weinberg 1974, 1984, 1990, 1992). When compared with bone marrow aspirations, serum ferritin values

correctly detected iron deficiency in 90 percent of patients; serum iron in 41 percent; total iron binding capacity in 84 percent; and transferrin saturation in 50 percent (Burns et al. 1990). Erythrocyte sedimentation rate (ESR) is a nonspecific measure of the presence of infection or inflammation. If used in combination with serum ferritin, ESR can be useful in distinguishing anemia of chronic disease from iron deficiency anemia (Beganovic 1987; Charache et al. 1987; Witte et al. 1988).

Although diet is most often attributed as the cause of iron-deficiency anemia, in the absence of blood loss or parasites, even the most frugal diets rarely result in iron-deficiency anemia. According to C. K. Arthur and J. P. Isbister (1987: 173), "iron deficiency is almost never due to dietary deficiency in an adult in our community [i.e., Western society]." The average person requires very little iron intake to replace that lost through perspiration, menstruation, and urination because as much as 90 percent of the iron needed for the formation of new blood cells is derived from recycling senescent red blood cells (Hoffbrand 1981). In addition, individuals suffering from iron deficiency absorb significantly more iron from the same diet and excrete significantly less iron than nondeficient individuals. In other words, "the intestine can adjust its avidity to match the body's requirement" (O'Neil-Cutting and Crosby 1987: 491). Nonetheless, it has been estimated that approximately 30 percent of a world population of nearly 4.5 billion are anemic, with at least half of this 30 percent (500 to 600 million people) thought to have iron-deficiency anemia (Cook and Lynch 1986). Even in affluent Western societies, 20 percent of menstruating females have been reported to be iron deficient (Arthur and Isbister 1987: 172). Recent surveys in the United States, however, indicate that these figures are much too high (Cook et al. 1986).

Elsewhere, researchers have discovered, to their surprise, that diet is not always responsible for the acknowledged high morbidity of a particular region. According to P. Aaby (1988: 290-1), "We started out by assuming that malnutrition was the major determinant of high child mortality and that changing this pattern would have a beneficial effect on mortality. Much to our distress, we found very little evidence that nutritional status [as measured by the presence of anemia] was strongly related to the variation in nutritional practices."

Failure to distinguish anemia etiology correctly can be seen in several studies. For example, Arthur and Isbister report that "of 29 patients 'diagnosed' as having iron deficiency anaemia, only 11 patients in fact had true iron deficiency when reviewed by the authors. Most patients had the anaemia of chronic disease that was misdiagnosed as iron deficiency. There is a strong clinical impression amongst hematologists that this problem of misdiagnosis is not unique to hospital based specialists but also applies more

widely" (Arthur and Isbister 1987: 172). Almost 50 percent of infants diagnosed as iron deficient were found to have occult gastrointestinal blood loss without anatomic lesions, which was the source of their anemia (Fairbanks and Beutler 1988: 208). Although the following discussions are hampered somewhat by the fact that not all studies include serum ferritin tests needed to distinguish dietary iron-deficiency anemia from the anemia of chronic disease, an attempt is made to differentiate when possible and to indicate when this is not possible.

Diet and Iron-Deficiency Anemia

When iron-deficiency anemia is caused by severely deficient diets, it appears to be a straightforward problem: Not enough iron is ingested to replace losses. In reality, many cases of iron-deficiency anemia involve unidentified blood loss that causes, accentuates, and/or perpetuates dietary-induced iron-deficiency anemia. (Note that throughout this paper, we are specifically discussing iron and diet, not protein, vitamins, and other nutrients that are also integral parts of a diet).

Anemia is rarely caused by diet alone. This is partly because healthy humans lose only 1 to 3 milligrams (mg) iron per day and partly because almost all food and, in many areas, drinking water contain some iron. Women in a Swedish study lost between 0.6 and 0.7 mg of iron through menstruation and 95 percent lost on average *less* than 1.4 mg per day (Fairbanks and Beutler 1988: 205).

Most individuals are able to replenish these losses with a normal diet. For example, Westerners ingest meat almost daily and consume additional iron in fortified foods (such as that contained in many wheat products). Wine and cider can provide 2 to 16 mg or more iron per liter (Fairbanks and Beutler 1988). Moreover, depending on the quantity of the body's iron stores, between 15 and 35 percent of heme iron, which is found in red meat, poultry, and fish, and between 2 and 20 percent of nonheme iron in plants is absorbed (Monsen 1988: 787). Higher amounts of nonheme iron in vegetables are absorbed when consumed with heme iron in meat or in combination with iron absorption enhancers, such as ascorbic acid (Hallberg, Brune, and Rossander 1989). Several studies of the amount of iron ingested by North Americans indicate that normal individuals who are not suffering from unnatural blood loss consume sufficient iron to prevent deficiency. For example, a study of 51 women found that those who consumed only poultry or fish or who were vegetarians had a mean iron intake of 12 mg, whereas those who regularly ate red meat had a mean iron intake of 13 mg (Monsen 1988: 789). The mean iron intake of 162 women in another study was 11.1 mg (Subar and Bowering 1988). Female long-distance runners ingested between 11.5 mg and 14 mg per day (Manore et al. 1989).

Blood loss and true iron deficiency also result in increased iron absorption. Research shows that iron-replete men with sufficient iron stores absorb 2.5 percent of nonheme iron ingested, as compared to 26 percent of the heme iron consumed (Cook 1990). These percentages can be contrasted with iron-deficient males who absorb 22 percent more nonheme iron than normal males and 47 percent more heme iron (Cook 1990: 304). The same flexibility applies to iron loss. Normal males lose about 0.9 mg of iron per day. In contrast, hypoferremic males lose only about 0.5 mg per day, and hyperferremic males lose about 2.0 mg per day (Finch 1989).

Although 82 percent of women studied in New York State recorded an average consumption of only 75 percent of the United States Recommended Dietary Allowance (RDA) of iron, the amount ingested was a mean of 11.1 mg, which exceeds the amount of iron loss through normal physiological processes (Subar and Bowering 1988; also see Brennan et al. 1983 for similar conclusions based on a study of low-income pregnant women). Thus, if we arbitrarily assume an absorption rate of 18 percent of all iron consumed, the mean iron available is well above the 1.4 mg required by menstruating women. This suggests that previously recommended replacement needs of average menstruating women may have been set too high. More scientists today recommend that average menstruating women need to absorb only 1.4 mg iron daily to replace losses (Monsen 1988: 786). In fact, the RDA is higher than the average iron loss for all groups. As a result, many people take in less than the RDA levels of iron without suffering dietary iron deficiency.

Albeit rare, severely iron-deficient diets do exist, though not necessarily because of low iron intake. Adult East Indian vegetarians ingest large quantities of iron, as much as 30 mg per day (Bindra and Gibson 1986). At the same time, they consume such large amounts of inhibitory substances that some suffer from iron-deficiency anemia (Bindra and Gibson 1986). These items include tannin (especially from tea); dietary fiber; coffee; calcium phosphate; soy protein; and phytates in bran and nuts, such as walnuts, almonds, peanuts, hazelnuts, and Brazil nuts (Macfarlane et al. 1988; Monsen 1988; Muñoz et al. 1988). The larger the amount of phytate consumed, the less iron is absorbed (Hallberg, Brune, and Rossander 1989). What appears to occur is that although the diet contains sufficient iron for replacement of normal physiological losses, inhibitory factors limit the absorption of that iron.

In addition to malabsorption of iron, moderate-to-heavy consumption of fish has been associated with a significantly increased bleeding time test in both males and females (Herold and Kinsella 1986; Houwelingen et al. 1987; Sullivan 1989). The prolonged bleeding time presumably results in significantly more blood loss and lower iron stores. This

may explain why young women who habitually eat fish as their major source of protein have levels of serum ferritin and anemia similar to those of strict vegetarians who consume iron inhibitory substances (Worthington-Roberts, Breskin, and Monsen 1988). Here again, we see that the amount of iron ingested is not directly related to levels of anemia. Occult blood loss and iron absorption are more important as relevant factors fostering iron-deficiency anemia than is the actual amount of dietary iron ingested.

Illness and the Anemia of Chronic Disease

The anemia of chronic disease/inflammation has the same hematological presentation as dietary iron deficiency anemia, with the exception of normal to elevated serum ferritin levels. As stated earlier, both normal-to-elevated serum ferritin (reflecting adequate or above-normal iron stores) and elevated ESR indicate the presence of infection and/or the inflammatory process. In fact, such conditions, created by the body's generalized nonspecific defense against invading pathogens or neoplasia (cancer), reduce the availability of circulating iron. This denies to many bacteria, parasites, and neoplastic cells the iron they need to proliferate. These hematological changes are caused by the production of interleukin-1, interleukin-6, and tumor necrosis factor alpha (Weinberg 1992).

There is abundant research to support the conclusion that anemia is one of the body's defenses against disease (fever and the manufacture of various cytokines are other defenses [Kent, Weinberg and Stuart-Macadam 1994]). A broad spectrum of illnesses is associated with the anemia of chronic disease (Cash and Sears 1989). The role of iron in enhancing infections and anemia as a defense have been detailed in a number of publications by Eugene Weinberg (1974, 1984, 1986, 1990, 1992). Most bacteria and parasites require iron but cannot store it. They survive and thrive by extracting needed iron from their host. As a consequence, increased available iron can be detrimental to health, as can be seen in modern populations receiving iron supplements or consuming large amounts of dietary iron. For example, the incidence of serious bacterial infections increased significantly when infants were given intramuscular iron-dextran (Barry and Reeve 1977: 376):¹ "When the iron administration was stopped the incidence of disease in Polynesians decreased from 17 per 1,000 to 2.7 per 1,000 total births." Other research substantiated the association between an increased incidence of bacterial infections, such as *Escherichia coli* sepsis, and the administration of intramuscular iron-dextran in neonates (Becroft, Dix, and Farmer 1977).

A number of health problems are connected with increased iron intake. In New Guinea, infants with respiratory infections given iron-dextran had longer hospital stays and higher morbidity than those not given iron supplements (Oppenheimer et al. 1986a). Malarial infection was greatly increased in infants with

higher transferrin saturation levels, leading the investigators to conclude that hypoferrremia has a protective role against malaria (Oppenheimer et al. 1986b). Furthermore, in a study of 110 patients in Africa, those with anemia had fewer malarial attacks than those with higher iron levels (Masawe, Muindi, and Swai 1974). Anemic patients who did not exhibit signs of malaria developed malaria *after* iron therapy was initiated. In contrast to nonanemic Turkana from Kenya, mildly anemic Turkana who consume little meat, poultry, or other heme iron have a lower incidence of infectious diseases, including such various infections as malaria, brucellosis, and amebiasis (Murray, Murray, and Murray 1980a). Among the Maasai, anemic individuals had a significantly lower incidence of amebic dysentery. Examination of the cow's milk consumed by Maasai "showed that it not only had a concentration of iron below the minimum necessary for the growth of *E. histolytica* [an amoeba] but also contained partly saturated lactoferrin and transferrin, which may actively compete with the parasite in the colon for ambient iron" (Murray, Murray, and Murray 1980b: 1351).

When associated with malaria and other infections, serum ferritin levels are generally increased in anemic children (Adelekan and Thurnham 1990). This indicates that iron stores are present in these children, despite the lowered serum iron, transferrin saturation, and hemoglobin levels, all of which suggests strongly that the anemia is not the result of dietary inadequacies. The body removes circulating iron to storage in an attempt to reduce its availability to invading malaria parasites or bacteria. Food taboos that restrict young children's intake of iron by prohibiting the consumption of meat are found in many areas where malaria is endemic, from Africa to India to Malaysia and New Guinea (Lepowsky 1985). These taboos may be partially explained as an attempt to aid the body's hypoferrremic defense against malaria (Lepowsky 1985: 120).

Some diseases result in highly elevated transferrin saturation levels that make iron much more accessible to pathogens. In diseases such as leukemia, patients have been observed with sera that were 96 to 100 percent saturated (normal is 16 to 50 percent). As a consequence, leukemia patients are unusually susceptible to infection (Kluger and Bullen 1987). Thus in one study of 161 leukemia patients who died, 78 percent of the deaths were attributed to infection and not to leukemia directly (Kluger and Bullen 1987: 258).

Another investigation provides an excellent demonstration of the direct relationship between anemia of chronic disease and infection/inflammations. Over a period of 30 days, researchers analyzed the blood of healthy, well-nourished, and nonanemic infants before and after they were immunized with live measles virus (Olivares et al. 1989). After immunization, the hemoglobin, serum iron, and transferrin

saturation levels fell significantly, whereas serum ferritin levels rose significantly as the body shifted circulating iron to storage. These levels persisted for 14 to 30 days, while the body produced hypoferrremia, or low levels of circulating iron, in an attempt to make iron less available. Even "in those infants with prior evidence of normal iron status, the viral process induced changes indistinguishable from iron deficiency" (Olivares et al. 1989: 855). The authors noted that changes in the white blood cells mimicked a bacterial infection, as did changes in iron levels; in both cases, rapid proliferation of the virus imitated the rapid proliferation of bacteria (Olivares et al. 1989: 855). The body was unable to differentiate between the proliferation of bacteria and the proliferation of the virus and responded in the same manner by producing hypoferrremia.

Anemia in Prehistory

In prehistoric populations, the distribution of anemia equaled or surpassed that found in today's world. There was a general increase in anemia through time, ranging from a few, extremely rare occurrences in the Paleolithic (Anderson 1968) to the Mesolithic (Janssens 1970), to more common occurrence in the Neolithic (Angel 1967), and the Bronze Age (Cule and Evans 1968). Anemia was even more frequent in recent groups, such as the prehistoric American southwestern Pueblos (El-Najjar 1977).

Prehistoric study is possible because anemia is identifiable in skeletal material by a cranial deformation called porotic hyperostosis (also known as cribra orbitalia when the orbits are affected). Physical effects include cranial lesions that are characterized by a sievelike porosity involving parts of the outer skull, often causing a thinning of the outer table, textural changes, and sometimes a "hair on end" appearance (Moseley 1965; Stuart-Macadam 1987). Porotic hyperostosis was originally thought to result from hereditary anemias, such as thalassemia or sickle cell anemia. More recently, chronic acquired anemia has been identified from roentgenograms (X rays) as a cause of porotic hyperostosis visible in living young anemic children (Shahidi and Diamond 1960; Powell, Weens, and Wenger 1965). More recent and conclusive studies confirm the link between porotic hyperostosis and acquired anemia (Stuart-Macadam 1987).

Whereas it is not difficult to identify anemia in skeletal populations, it is most difficult to differentiate dietary iron-deficiency anemia from the anemia of chronic disease. At first it was thought that the chemical makeup of bones was an indirect measure of the amount of meat consumed (Bumsted 1980, 1985). However, problems of chemical contamination through mineral loss and contact with minerals in soil have yet to be resolved. A study specifically designed to test the reliability of current bone chemical analyses concluded that their "results suggest that post-

mortem alteration of dietary tracers in the inorganic phases of bone may be a problem at all archaeological sites and must be evaluated in each case" (Nelson et al. 1986: 1941). Other problems associated with stable isotope and trace element analyses of bone are numerous, and interpretations utilizing current techniques are not conclusive (Klepinger 1984; Aufderheide 1989; Keegan 1989; Sillen, Sealy, and van der Merwe 1989). However, indirect measures of dietary iron and disease, which rely on a basic understanding of iron, anemia, diet, and disease, are possible. The presence of porotic hyperostosis in nonhuman primates, such as chimpanzees, gorillas, orangutans, and various species of Old World monkeys, including baboons (Hengen 1971), reinforce the unlikelihood that an impoverished diet is frequently the cause.

Anemia in the New World

Poor diet, sometimes in combination with infectious diseases, has served as the usual explanation for skeletal evidence of anemia among prehistoric Native Americans from the eastern United States to South America. However, faunal remains, mobility patterns, new medical information, and other nontraditional lines of evidence indicate that disease, rather than diet, may often be a better interpretation.

Eastern United States. Although some investigators still claim a maize-dependent diet as the sole or primary cause of porotic hyperostosis in prehistoric southeastern skeletal populations (Robbins 1986), more recent reevaluations suggest that diet may *not* have been a major cause. During the Mississippian period, an increase in porotic hyperostosis in children, particularly those under the age of six, occurred in the lower Illinois Valley. Nutritional anemia alone was probably rare in this age group, especially among breast-fed children. However, the condition coincides with a general increase in destructive bone lesions indicative of chronic inflammations in skeletal populations (Cook 1984: 257-9). During the period of these burials (between A.D. 1100 and 1300), Cahokia grew to be the largest pre-Columbian city in North America, with a population of 25,000 to 43,000 (Cook 1984). Non-Western sedentary, aggregated communities often have heavy pathogen loads, as discussed in the section "Anemia of Sedentism." The skeletal populations exhibit a concomitant increase in anemia and inflammation. This combination illustrates the link between increased population density and increased health problems. Diet cannot be implicated in anemia because during this time the:

Mississippian diet based on archaeological data and ethnohistoric documentation . . . seems remarkably abundant in the essential elements: protein from both animal and vegetable sources, carbohydrates from cultivated and collected plant foods, and oils from seeds, nuts,

and animal fat, all of these rich in minerals and vitamins from undepleted soils. . . . This ecological diversity, coupled with the sophisticated systems of multiple cropping, surplus food storage, and redistribution described by early European observers, offered protection against nutritional stress from local fluctuations in weather and resource populations. (Powell 1988: 58)

Also during the Mississippian period at Dickson Mounds, Illinois, the incidence of porotic hyperostosis rose from 13.6 percent to 51.5 percent, which coincided with dramatic increases in infectious bone lesions (Goodman et al. 1984: 289). According to investigators, samples from earlier periods had infection rates of 25 percent of the tibiae examined, but then rose to 77 to 84 percent during the later Mississippian period (Goodman et al. 1984). Although the authors attributed the rise in infectious lesions to nutritional stress brought on by an intensification of agriculture in the later period, a change in demography and, particularly, in aggregation may well be responsible. Studies indicate that the earlier Late Woodland sites in the area had population estimates of 50 to 75 individuals; later Mississippian sites had 440 to 1170 individuals (Goodman et al. 1984). Such an important and dramatic demographic change must have had serious health consequences. These Illinois examples are not isolated cases for the eastern United States. In fact, similar rates of increased infection and porotic hyperostosis through time are reported for the Ohio River Valley (Perzigian, Tench, and Braun 1984) and Georgia (Larsen 1984).

Southwestern United States. A maize-dependent diet has also been implicated in the increase through time of porotic hyperostosis in skeletal material from the southwestern part of the United States. Mahmoud Y. El-Najjar and colleagues were among the first and certainly the most prolific to detail the association between frequency of porotic hyperostosis over time and increasing horticultural activities (El-Najjar et al. 1976; El-Najjar 1977). However, as in the eastern United States, the archaeological data from the Southwest do not seem to support a view of diets sufficiently impoverished as to account for the spectacular increase in porotic hyperostosis through time in Anasazi skeletal populations. Instead, archaeological, ethnographic, and ethnohistorical sources all indicate a varied diet, including meat from both wild and domesticated animals, with the latter including turkeys and dogs (Kent and Lee 1992). In addition, wild plants were consumed in combination with maize, beans, and squash, all of which constituted an adequate diet in terms of iron (Kent 1986).

What, then, caused the increase in porotic hyperostosis? As the Anasazi began to adopt horticulture in general, and maize cultivation in particular, they became more sedentary and aggregated, living in

communities called pueblos. Although this was not a linear progression that occurred everywhere, neither was the rise in porotic hyperostosis linear. Higher frequencies of porotic hyperostosis occur in skeletal populations from large, sedentary, aggregated pueblos in which disease was more common than in skeletal populations from smaller, more dispersed settlements where it would have been more difficult for disease vectors to infect populations (Kent 1986). Coprolite data (preserved fecal material) support this interpretation by documenting a lower parasite load for the upland populations in contrast to the valley groups. Coprolites containing parasites from various Anasazi sites are significantly correlated with porotic hyperostosis, whereas those containing different portions of either maize or animal residue are not (Reinhard 1992). In other words, dietary evidence of reliance on meat or maize was not correlated with skeletal evidence of anemia, but parasitism was correlated with porotic hyperostosis.

Elsewhere in the United States. Anemia, apparently caused by chronic diseases and inflammations, produced relatively high levels of porotic hyperostosis throughout the prehistoric United States. It can be seen in coastal California Native American skeletal populations who had a heavy dietary dependence on marine resources (Walker 1986). Eskimo skeletal populations also reveal relatively high percentages of porotic hyperostosis, some of which are even higher than those of southwestern and eastern prehistoric populations (Nathan 1966). Certainly the Eskimo example of porotic hyperostosis was not the result of an insufficient intake of meat. Instead, it was probably the result of seasonally sedentary aggregations and associated disease common to the winter villages. Prehistoric skeletons from all parts of Texas exhibit much lower incidences of porotic hyperostosis; less than 5 percent of 348 adult crania and 15.1 percent of 73 juvenile crania had any evidence of porotic hyperostosis (Goldstein 1957). Although precise locations of the various skeletal material were not cited, it is probable that the populations examined represent the nomadic groups that inhabited much of Texas; if so, this would explain the low incidence of infectious disease and porotic hyperostosis.

Central and South America. High frequencies of porotic hyperostosis were found on the crania of prehistoric Maya from Central America. For example, 21 skulls of Mayan children 6 to 12 years old were found in the Sacred Cenote of Chichén Itzá in the Yucatan; 67 percent had porotic hyperostosis. A number of the adult crania from the cenote also had healed examples of the pathology (Hooton 1940). Prehistoric Maya living in Guatemala suffered from both a high rate of infectious diseases, as evidenced by lesions thought to be related to yaws or syphilis, and to dietary inadequacies, such as vitamin C deficiency, as

evidenced by bone lesions (Saul 1973). As a consequence, it is difficult to determine whether the porotic hyperostosis was caused by true dietary deficiency, by infectious diseases, or by a combination of the two.

In contrast, a study of Ecuadoran skeletal populations clearly demonstrates the influence of disease in promoting porotic hyperostosis. Although coastal populations ate iron-rich seafood, in addition to meat and cultivated products, skeletal populations reveal increasingly higher frequencies of porotic hyperostosis over time as they became more sedentary and aggregated (Ubelaker 1992). Coastal peoples suffered from conditions aggravated by large-village life, such as probable parasitic infestations, including hookworm, which causes gastrointestinal bleeding.² An examination of mummies from coastal Peru and Chile indicates that the most common cause of death was acute respiratory disease in adults and children (Allison 1984). Approximately half of the mummified individuals died from their first attack of pneumonia (Allison 1984). In contrast, porotic hyperostosis is not common in skeletons of maize farmers who occupied more dispersed habitations in the highlands of Ecuador, where parasites cannot survive because of the altitude and cold climate (Ubelaker 1992).

Although prehistoric cases of dietary-induced porotic hyperostosis may occur, they probably are not representative of entire populations. Iron-deficiency anemia caused by extremely inadequate diets is thought to result from factors not common until the twentieth century. Although not all paleopathologists (particularly those who are more diet-oriented and are less hematologically oriented), agree with the emphasis placed here on the anemia of chronic disease, many recent studies and reinterpretations agree that this is a better explanation than iron deficiency.

Anemia in the Old World

The Old World represents a vast area occupied by humans for a long time period, and thus only a few examples of prehistoric incidences of anemia are presented here.

Hereditary (for example, thalassemia) and acquired cases of anemia from the Mediterranean area cannot be differentiated, although porotic hyperostosis is found in skeletal populations from the area (Angel 1984). In the Levant, Mesolithic Natufians apparently had good health, as did the Paleolithic hunter-gatherers who preceded them, although sample size is very small (Smith et al. 1984). They remained a relatively healthy population during the early Neolithic. Deterioration began in later periods when population aggregations grew larger and more permanent and infectious diseases became endemic. According to P. Smith, O. Bar-Yosef, and A. Sillen, "This deterioration . . . seems to be related to chronic disease rather than to periodic

bouts of food shortages, as indicated by the distribution of developmental lesions in the teeth and bones and the poor condition of all individuals examined" (Smith et al. 1984: 129).

In South Asia, there was a similar trend toward increasing porotic hyperostosis and infection, which has been attributed to nutritional deficiencies and disease (Kennedy 1984) but may also have been caused by demographic factors. Skeletons from the Harappan civilization city of Mohenjo-Daro on the Indus River, Pakistan, have a relatively high incidence of porotic hyperostosis; this led one anthropologist to suggest the possible presence of thalassemia and malaria in the area 4,000 years ago (Kennedy 1984: 183). However, the occurrence of malaria is difficult to evaluate, particularly in light of the high population density of Mohenjo-Daro, which could have promoted a number of infectious diseases common to such sedentary aggregations.

Porotic hyperostosis is found at varying frequencies throughout European prehistory (Hengen 1971), but as no systematic temporal study has been done to date, it is not possible to delineate and interpret trends. However, an interesting geographical pattern has been discerned that suggests that populations located closer to the equator have higher levels of porotic hyperostosis. This might be the result of generally higher pathogen loads in these areas (Stuart-Macadam 1992). However, porotic hyperostosis was well represented, particularly in children, in a large Roman-British cemetery (Stuart-Macadam 1982). This is interesting because it has been suggested that lead poisoning, which can cause severe anemia, was one of the factors that contributed to the ultimate collapse of the Roman Empire (Gilfillan 1965).

According to some research, prehistoric Nubian populations experienced nutritional deficiencies after they became agriculturalists (Armelagos et al. 1984). This interpretation is based on the occurrence of porotic hyperostosis, long-bone growth patterns compared to North Americans, dentition abnormalities, and premature osteoporosis. Moreover, bone growth patterns of skeletons from a medieval Nubian Christian cemetery, (A.D. 550 to 1450) have been used to interpret nutritional stress (Hummert 1983). Nonetheless, it is usually recognized that such stress can be caused by a number of nutritional deficiencies not necessarily related to iron. More recent coprolite data have been employed in support of the hypothesis that an impoverished diet caused the high frequency of porotic hyperostosis in Christian Nubian populations (Cummings 1989). However, it is possible to interpret the data quite differently. For example, it should be noted that meat does not usually preserve in coprolites, and humans do not usually ingest large bones that would be indicative of meat consumption. Therefore, the amount of meat consumed may be substantially underrepresented when interpreting coprolite data. Bearing this in mind, it is impressive that

33.3 percent of the 48 coprolites analyzed contained evidence of fish bones or scales, as well as one pig bone and an unidentifiable mammal bone (Cummings 1989: 86–92). Such evidence contradicts the contention that a diet poor in iron produced the high incidence of porotic hyperostosis in this population.

Anemia in Australia

Frequencies of both infection and anemia (porotic hyperostosis) are low in the desert region of Australia, particularly when compared to other parts of the continent, such as the Murray River area (Webb 1989). Murray River skeletons display a pronounced increase in porotic hyperostosis and infections coinciding with archaeological evidence of restricted mobility, aggregation, and possible overcrowding (Webb 1989: 145–8). Such evidence again suggests that the anemia in question is the result of chronic disease and not diet. Though lower than in the Murray River area, the relative high incidence of anemia and infection that occurred among prehistoric Aborigines occupying the tropical portions of Australia can be attributed to parasitism, such as hookworm (Webb 1989: 155–6).

Acquired Anemia in Today's World

The perspective of dietary deficiency and the anemia of chronic disease presented here permits new insights into a variety of issues facing contemporary populations. Because iron-deficiency anemia and the anemia of chronic disease have not always been distinguished, particularly in earlier studies, the two are discussed together and are differentiated whenever possible.

Anemia of Sedentism

As noted, heavy pathogen loads are characteristic of non-Western sedentary aggregations. In such settings, high morbidity rates are visible, not only hematologically but also in the frequency of parasitic infections. For example, studies indicate that nomadic Amazonians had a lower frequency of parasitic infection than seminomadic horticulturalists and sedentary villagers. In these latter populations, 100 percent of some age groups were infected with parasites (Lawrence et al. 1980). As a whole, the sedentary populations had many more multiple parasitic infections, ranging from 4.2 to 6.8 species per person (Lawrence et al. 1980). Sedentary and aggregated Brazilian Amazonian villages had roundworm (*Ascaris*) infections, ranging from 65 to 100 percent of the population, and heavy worm burdens, including hookworm and whipworm (*Necator americanus*, *Trichuris trichura*) (Chernela and Thatcher 1989). This situation contrasts with the inhabitants of more nomadic and dispersed villages that had half that rate of roundworm infection (34 percent) and light worm burdens that were asymptomatic (Chernela and Thatcher 1989).

From 60 to 100 percent of the residents of two Colombian villages were found to be infested with parasites (Schwaner and Dixon 1974). In one that lacked sanitation of any kind and in which shoes were rarely worn, 100 percent of the population had either double or triple infections (roundworm, whipworm, and hookworm). Of the 60 percent of those infected in the other village (that had outdoor latrines and where shoes were worn), 70 percent were infected by more than one species (Schwaner and Dixon 1974: 34). Sedentism, aggregation, and the lack of adequate sanitation are all implicated in creating breeding grounds for parasitic and other types of infections. Such conditions lead to high morbidity and chronic hypoferrremia as the body attempts to defend itself against the heavy pathogen load. Furthermore, heavy infestations of hookworm and other parasites cause blood loss that even good diets cannot replenish.

The relationship between sedentism and anemia can be more easily seen by comparing nomadic and sedentary people who consume similar diets, thereby factoring out diet as a causal variable of anemia. Hematological research among recently sedentary and still nomadic Basarwa ("Bushmen" or San, as they have been referred to in the literature) illustrates hypoferrremia operating in a modern population. The 1969 population of the !Kung Basarwa (who live in the northwestern part of the Kalahari Desert) contained a few ill individuals, as do most populations, but was judged as mostly healthy (Metz, Hart, and Harpending 1971). Their meat intake did not change dramatically between the dry seasons of 1969 and 1987 (Kent and Lee 1992). There were, however, significant changes in mobility patterns, from a relatively nomadic to a relatively sedentary pattern. Concomitant higher morbidity rates ensued. Also, the number of individuals with below-normal serum iron and transferrin saturation levels rose significantly between 1969 and 1987 (Kent and Lee 1992). It is significant that the 1987 Dobe !Kung population, with a roughly adequate meat intake, had more individuals with subnormal serum iron and transferrin saturation values than did the Chum!kwe !Kung population, with an acknowledged deficient diet (Fernandes-Costa et al. 1984; Kent and Lee 1992). It is furthermore significant that no individuals had low serum ferritin values among the Dobe !Kung, which would indicate a dietary deficiency.

To evaluate the role of sedentism in promoting the anemia of chronic disease, a hematological study was conducted in 1988 among a different group of recently sedentary Basarwa living near the Kutse Game Reserve in the eastern half of the Kalahari (Kent and Dunn 1993). Their diet was repetitive but adequate; meat was consumed several times a week (Kent and Dunn 1993). The Kutse Basarwa often complained of illness; respiratory problems were com-

mon. Both serum iron and transferrin saturation means were significantly lower for the 1988 Kutse Basarwa than those obtained in 1969 or 1987 among the Dobe Basarwa (Kent and Dunn 1993). At the same time, serum ferritin levels were higher in the adult Kutse population than in the 1987 adult !Kung. This is consistent with the view that the anemia of chronic disease is more prevalent in situations of high morbidity where the body attempts to protect itself from continuous cycles of insult. The children's mean ferritin remained approximately the same (Kent and Dunn 1993).

If hypoferrremia operates in response to heavy and chronic pathogen loads, then it should be visible in more than just the hematological profile of a population. Ethnographic observations and informant interviews indicate that there is a high level of morbidity at Kutse but not a deficient diet (Kent and Dunn 1993). Both the 1987 !Kung and 1988 Kutse Basarwa hematological studies indicate that the anemia of chronic disease is common in these populations and is activated in situations of high morbidity. A 1989 follow-up hematological study of the Kutse community reinforces the interpretation of an adequate diet but high morbidity as a result of aggregation (Kent and Dunn n.d.). As pointed out, anemia is not a body defect or disease but is actually a defense against a heavy pathogen load, which operates by reducing circulating levels of iron required by organisms to proliferate.

Anemia of Development and Modernization

Development and modernization projects are often promoted both by local governments and by international businesses and foreign governments as a means to westernize a developing country. Generally, nontraditional subsistence activities are encouraged because the government or politically dominant group believes the change to be beneficial. Extremely deficient diets documented by Susan Stonich (1991), Benjamin Orlove (1987), and others are the consequence of governmental pressure to grow export crops for a world market, which is a relatively recent phenomenon. This type of severe dietary deprivation has usually been correlated with colonialism (Franke 1987; Ross 1987).

Nomadic peoples have been targets of development and modernization schemes. For example, sedentarization at Chum!kwe, located in Namibia on the western edge of the Kalahari Desert, began in 1960. At that time the South African administration of the Namibia Territory decided to move all the !Kung Basarwa of Nyae Nyae into a single settlement (Kent and Lee 1992). For the next 20 years the waterhole at Chum!kwe (augmented by boreholes) supported a standing population of 800 to 1,000 people. These conditions of high aggregation, with its attendant social, nutritional, and alcohol-related stresses, were prevailing when a 1981 study was made (Fernandes-Costa et al. 1984).

At Chum!kwe in 1981, hunting and gathering had fallen to very low levels due to the high level of aggregation and reduced mobility of the population (Fernandes-Costa et al. 1984). Store-bought foods and government rations, high in refined carbohydrates, formed the mainstay of the diet; some milk and grain from home-grown sources supplemented the diet. Meat was eaten infrequently. By the 1980s alcohol abuse had become a major social problem, with Saturday night brawls often resulting in injury and death (Volkmann 1982). Young children as well as adults drank beer, which sometimes provided the only calories consumed all day (Marshall and Ritchie 1984: 58). The health of the Chum!kwe population was portrayed as poor.

The investigators attributed the incidence of anemia primarily to nutritional causes. However, 35.9 percent of all adults had a serum ferritin value above 100 nanograms per milliliter (ng/mL), indicating the presence of the anemia of chronic disease. No men and only 6 percent of women had subnormal serum ferritin levels (Fernandes-Costa et al. 1984). Diet was also thought to be responsible for the 33 percent of children who had subnormal serum ferritin levels (less than 12 ng/mL), although bleeding from parasitic infections could be another explanation. In fact, 3 of 40 stool samples contained *Giardia lamblia* and 8 of 40 samples had hookworm (*Necator americanus*) (Fernandes-Costa et al. 1984: 1302). Both parasites can cause lowered serum ferritin and other iron indices through blood loss and competition with the host for nutrients. Unfortunately, the age and sex of the afflicted are not mentioned; however, children tend to become infected from parasites more often than adults and tend to have a heavier parasite load than adults. Thus, parasites might explain the number of children with subnormal serum ferritin values.

A very similar situation occurred among the Australian Aborigines. They were encouraged to live at settlements where they subsisted on government subsidies of flour and sugar (Taylor 1977). The consequence of this planned social change implemented by the government was malnutrition and frequent infections:

After weaning from breast milk, the children graduated to the staple and nutritionally deficient diet of the settlement and entered its unhygienic, usually overcrowded, and certainly highly pathogenic environment. Here they acquired repeated respiratory and gastrointestinal infections which, if they did not prove fatal, may have impaired the children's ability to absorb nutrients from the diet [through malabsorption caused by the anemia of chronic disease and by numerous types of infections]. Thus the problem was compounded and a vicious

cycle seemed to be set up in which the survivors of the process proceeded to adulthood to rear more children under the same circumstances. (Taylor 1977: 147–8)

Anemia around the World

In the United States the frequency of anemia is fairly low, approximately 6 percent or less for children under 17 years of age; more girls than boys have subnormal levels (Dallman 1987). The number of children with subnormal hemoglobin values declined from 7.8 percent in 1975 to 2.9 percent in 1985 (Centers for Disease Control 1986). A similar decline occurred in adults as indicated in the Second National Health and Nutrition Examination Survey (NHANES II, 1976–80). In this study of 15,093 subjects, anemia was found in 5.7 percent of infants, 5.9 percent of teenage girls, 5.8 percent of young women, and 4.4 percent of elderly men (Dallman, Yip, and Johnson 1984). The later surveys included serum ferritin measurements and differentiated between iron-deficiency anemia and anemia of chronic disease (Expert Scientific Working Group 1985). Therefore, the decline may be more illusory than real in that the more recent surveys separated the two anemias, which made the overall frequency of each appear to be less. Although based on a small sample size, anemia of chronic disease accounted for 10 percent of anemic females between 10 to 44 years of age, 34 percent of anemic females between 45 and 74 years of age, and 50 percent of anemic males (Expert Scientific Working Group 1985: 1326).³

There are a few populations with a lower rate of anemia than that in the United States. One example is the Waorani Indian horticulturalist-hunters of eastern Ecuador. In 1976, none of the 114 individuals studied, including males and females of all ages, had subnormal hemoglobins or hematocrits, despite a few light helminth infestations and four children with presumed bacterial-induced diarrhea (Larrick et al. 1979: 155–7). Once again, the low frequency of anemia was probably the result of a seminomadic, dispersed settlement pattern.

Anemia frequency is much higher in other parts of the world. During the 1970s, studies conducted of nonwhite children in South Africa showed a higher prevalence of iron-deficiency anemia (38.8 percent) and anemia of chronic disease (18.8 percent) than figures reported for the same time period in the United States (Derman et al. 1978).

In regions such as southeast Asia, a high incidence of hookworm infections cause bleeding. Surveys in these areas show that the frequency of anemia is correspondingly high, ranging from 25 to 29 percent in males and from 7 to 45 percent in nonpregnant women and children (Charoenlarp et al. 1988). Approximately 20 percent of the subjects given oral iron for three months as part of a study in Thailand remained anemic, and some of the individuals ended up with serum ferritin levels above 100 µg/L

(Charoenlarp 1988: 284–5). Both the lack of response to oral iron supplementation and the serum ferritin levels above 100 are common indications of the anemia of chronic disease; the other incidences of anemia can be attributed to various amounts of hookworm and other parasitic infections. Between 5 to 15 percent of women, 3 to 27 percent of children, and 1 to 5 percent of men in Burma were anemic, but no serum ferritin values were measured; hookworm infection rates varied from 0 to 70 percent, depending upon locality (Aung-Thun-Batu et al. 1972).

Anemia was found in 22 percent of pregnant women, 12 percent of nonpregnant women, and 3 percent of men in seven Latin American countries (Cook et al. 1971). Unfortunately, serum ferritin was not measured, and so it is not possible to distinguish the percentage of individuals with iron deficiency or with anemia of chronic disease.

Anemia and parasites. Parasite load is an influential factor in the frequency of anemia around the world. Parasites both compete with the host for nutrients and can cause bleeding resulting in substantial iron losses. The body responds to parasitic infections with hypoferremia to deprive the parasites of needed iron. Most common in the tropical regions of the world, parasitic distribution is related to the ecological complexity and diversity that characterize such habitations (Dunn 1968; Stuart-Macadam 1992). Hookworms can cause as much as 0.21 ml of blood loss per worm per day in infected individuals (Ubelaker 1992).

In Bangladesh, where over 90 percent of children have helminth infections, between 74 percent and 82 percent of the children tested had subnormal hemoglobin levels (Florentino and Guirriec 1984). Serum ferritin studies were not included. However, because of the blood loss associated with helminth infections, individuals were probably suffering from iron-deficiency anemia, albeit not diet induced. Of 1,110 children tested in India, 33.8 to 69.4 percent were anemic; 26.7 to 65.3 percent suffered from roundworm (*Ascaris*) and *Giardia* infections, both of which cause anemia (Florentino and Guirriec 1984). Although the link with parasites is a little more tenuous among Nepal villagers, 67.6 percent of nonpregnant women had subnormal hemoglobin levels; the high consumption of dietary iron absorption inhibitors may also be a factor in causing anemia in this population (Malville 1987).

In Indonesia, 37.8 to 73 percent of children studied were anemic, and 22 to 93 percent were infested with hookworm, probably causing their anemia (Florentino and Guirriec 1984: 85). This study also reported that in Malaysia and Singapore, 83 percent of 30 children were anemic; in the Philippines, 21.1 to 47.2 percent of 2,509 children, depending on their age group, were anemic; and in China, 23 percent of 1,148 children were anemic.

Among various Pygmy hunter-gatherer groups, parasite levels are extremely high because of fecal contamination and the high level of parasites characteristic of tropical environments (Pampiglione and Ricciardi 1986). Various types of malaria (*Plasmodium falciparum*, *Plasmodium malariae*, and *Plasmodium ovale*) were found in 18.1 to 59.4 percent of 1,188 people from four different groups of Pygmies living in the Central African Republic, Cameroon, and Zaire (Pampiglione and Ricciardi 1986). Numerous other parasites were also present: hookworms (41.3 to 85.8 percent); roundworms (16.7 to 64.9 percent); amoebas (*Entamoeba*, 6.4 to 35.8 percent); *Giardia* (5.3 to 11.4 percent); whipworm (*Trichuris*, 77.9 to 91.9 percent); and others (Pampiglione and Ricciardi 1986). Many individuals suffered from several parasitic infections, in addition to yaws (10 percent with active symptoms), scabies, chiggers, and other afflictions (Pampiglione and Ricciardi 1986: 160-1). Neighboring Bantu-speaking farming peoples, despite more access to Western medical care and medicine, also have a high incidence of parasitism. Malaria (*P. falciparum*) was found in 91.2 percent of 321 persons; 26.8 percent suffered from amoebas; 53 percent from roundworms; 80.3 percent from hookworms; 69.8 percent from whipworm; and 16.5 percent from *Giardia* (Pampiglione and Ricciardi 1986: 163-4).

In Liberia, pregnant women in one study also had high rates of parasitism: 38 percent had hookworm; 74 percent had roundworm; and 80 percent had whipworm (Jackson and Jackson 1987). Multiple infections were also common; between 24 and 51 percent of the women had two infections and 12.5 percent had three infections (Jackson and Jackson 1987).

In Bolivia, 11.2 to 44 percent of children tested were anemic (Florentino and Guirriec 1984). Of these, 79 percent suffered from roundworm and 12 percent had hookworm infections (Florentino and Guirriec 1984).

The intention here is not to provide a detailed overview of anemia throughout the world, which would require an entire book. Rather, it is to show that anemia is a widespread condition in non-Western societies due to endemic chronic infections and parasitic diseases. In some of the cases mentioned in this section, poor health in general resulted from ill-planned development schemes or encouragement of cash or export crops at the expense of subsistence farming, creating contaminated water supplies and poor sanitation. This should present a challenge to Western societies to eradicate disease aggressively in developing countries. Development agencies need to become more aware of the underlying causes of anemia and high morbidity in many of these countries. Attempts to change agricultural patterns to produce more food per acre or the provision of iron supplements may not

achieve the goal of eliminating anemia on a worldwide basis.

Anemia in infants and children. The decline in the number of children with acquired anemia in Western nations is as impressive as the decline in the number of adults with anemia. One study in the United States showed that in 1971, 23 percent of 258 children were anemic, whereas in 1984, only 1 percent of 324 children were anemic (Dallman and Yip 1989). More recent surveys show that of those few children who are anemic, the majority have the anemia of chronic disease rather than dietary iron deficiency, as evidenced by normal-to-elevated serum ferritin levels (Reeves et al. 1984; Jansson, Kling, and Dallman 1986).

There is a vast literature on anemia in infants and children (for example, Pochedly and May 1987). Most relate the anemia to the rapid growth of children and the low iron content of diets, although many do not distinguish between iron-deficiency anemia and anemia of chronic disease (Ritchey 1987). For example, a study of 148 British toddlers claims that lower hemoglobin levels are associated with diet: (1) prolonged lactation and (2) early introduction of whole cow's milk (Mills 1990). Consumption of cow's milk before 12 months of age has been shown to cause gastrointestinal bleeding in 39.5 percent of infants fed cow's milk versus 9.3 percent of infants fed formula (Fomon et al. 1981).

We cannot assume, however, that all of these cases of childhood anemia were the result of dietary factors because serum ferritin was not measured. In fact, there are more compelling reasons to suggest that at least some of the anemia was the anemia of chronic disease. Women who breast-feed their children the longest (a year or more) in Western societies tend to be the underprivileged, who are also subjected to higher rates of infection or inflammation due to crowding, inadequate housing, and the stress of minority status (see the section "Anemia and Sociopolitical Class and Race"). Such children may be at higher risk from pathogens and have greater incidence of the anemia of chronic disease. Clearly, serum ferritin must be measured in order to make reliable conclusions from any study of anemia. It is interesting to note that a study of 9 hospitalized children (6 fed formula and 3 fed cow's milk) indicated that those with lower hemoglobin levels, attributed to starting cow's milk at 6 months of age, were less likely to be hospitalized for infections than infants fed formula who had higher hemoglobin levels (Tunnessen and Oski 1987). This could be interpreted as the effect of lower hemoglobin levels, which provides protection against infections both in the contracting of diseases and in their virulence. Subnormal serum ferritin values revealed in the 17.4 percent of infants fed cow's milk versus only 1 percent of infants fed enriched formula could be the result of increased diarrhea and gastrointestinal bleeding associated with cow's milk feeding, as well as to

allergies to cow's milk found in 0.3 to 7.5 percent of children (Foucard 1985; Tunnessen and Oski 1987).

Further difficulty in interpreting the etiology of anemia in children is caused by the fact that most minor illnesses that commonly afflict children, as well as immunization, can significantly depress hemoglobin levels (Dallman and Yip 1989). Iron-deficiency anemia is simulated in these cases because hemoglobin levels drop, but the anemia of chronic disease is the actual cause, as evidenced by normal serum ferritin levels. Moreover, serum iron levels drop significantly during the first year of life and are lower than adult levels as a normal physiological developmental change (Saarinen and Siimes 1977; Dallman and Reeves 1984).

In an attempt to combat anemia, most health-care workers routinely recommend iron-fortified infant formula without first ascertaining if an infant is anemic and if so, why. For example, the Women, Infant, and Children (WIC) program requires disadvantaged women to use iron-fortified formulas; nonfortified formulas are not provided (Kent, Weinberg, and Stuart-Macadam 1990). Fortified infant formulas were routinely prescribed by all but 16 percent of 251 physicians in Washington State (Taylor and Bergman 1989). It is hoped that the information presented here demonstrates the potential deleteriousness of this practice. Fortified formulas should be used only in cases where infants have anemia due to bleeding or other conditions known to create insufficient levels of iron to maintain health.

Anemia and Pregnancy

Pregnancy is often associated with a mild iron deficiency that has been linked to the nutritional needs associated with the rapid growth of a fetus. However, a few researchers have suggested that slight hypoferrinemia may help defend mother and fetus from invading pathogens (Weinberg 1987; Stuart-Macadam 1987). Such a defense may be particularly valuable during the latter phases of gestation when cell-mediated immunity lessens to prevent immunological rejection of the fetus (Weinberg 1987). That lower iron levels are a normal physiological part of pregnancy is supported by a number of studies that have failed to demonstrate any benefit derived from routinely prescribed prophylactic iron (Hemminki 1978; Bentley 1985). Nonetheless, many physicians continue to recommend that pregnant women take oral iron supplements.

There are anemic pregnant women with serum ferritin values below the normal 12 ng/mL. Premature labor was recorded in 48 percent of pregnant women with serum ferritin values below 10 ng/mL, in contrast to 11 percent who had normal serum ferritin levels (Goepel, Ulmer, and Neth 1988; Lieberman et al. 1988). However, a very large study of 35,423 pregnant women reported no reliable association between anemia and problem pregnancies:

When the hematocrits of women in term labor were compared with those of women in preterm labor, a spurious dose-response effect for anemia was created. We conclude that anemia is not a strong factor in the pathogenesis of preterm birth and that comparison of hematocrits from women who are in preterm and term labor produces biased results (Klebanoff et al 1989: 511).

Furthermore, and importantly, "we do not believe that there is sufficient evidence to justify a randomized clinical trial of treatment of borderline anemia during pregnancy" (Klebanoff et al. 1989: 515).

Anemia and the Elderly

Elderly men and postmenopausal women have the highest incidence of the anemia of chronic disease among all North Americans surveyed (Cook et al. 1986). Although some physicians have regarded anemia as a "normal" part of aging, a number of studies have shown that it often is the anemia of chronic disease resulting from increased vulnerability to pathogens and neoplasia, perhaps because of lowered immune defense (Zauber and Zauber 1987; Daly and Sobal 1989; Thomas et al. 1989). Serum ferritin levels rise with age, probably because the elderly have increased susceptibility to systemic insults (Zauber and Zauber 1987; Baldwin 1989; Daly 1989; Stander 1989). Whereas only 13 percent of 259 elderly persons had anemia, 77 percent had the anemia of chronic infections (Guyatt et al. 1990: 206). The frequency of infectious diseases may not be primarily the result of age alone but, instead, of social factors involved with aging in Western society, such as depression, poor care, and crowding in nursing homes. That is, the anemia of chronic disease may be more common among the elderly than the population in general because of social variables unique to this group.

Anemia and Sociopolitical Class and Race

Whatever the cause of anemia, dietary or chronic diseases, it is still the poor person's disease. While the prevalence of anemia has dropped for African-Americans as it has for all Americans, the frequency of anemia is still higher than among other groups. Anemia among African-Americans dropped from 21 percent between 1975 and 1977 to 19.2 percent between 1983 and 1985 (Yip et al. 1987). Unfortunately, serum ferritin was not measured in earlier surveys, making conclusions based on improved iron and general dietary nutrition problematic. However, later studies that include ESR (sedimentation rate, an indication of infection) show that the anemia of chronic disease is more prevalent among lower socioeconomic classes, including African-Americans (Yip and Dallman 1988).

It was once proposed that a racial characteristic of blacks is a significantly lower hemoglobin level than

that of whites, and as a result, some researchers suggested using separate hemoglobin standards based on race (Garn, Smith, and Clark 1975; Garn and Clark 1976; Garn, Shaw, and McCabe 1976). However, only hemoglobin and hematocrit were analyzed in these studies. Genetics, paleontology, physiology, anatomy, and other sciences all suggest that there are neither absolute races of humans nor corresponding genes that are restricted solely to one segment of the human population; that is, there are no "black" genes that all blacks share.⁴ So-called racial groups are based on continuous traits that are arbitrarily divided into supposedly discrete groups. This is particularly true in Western society where so much gene flow (that is, interracial matings) has occurred.

What, then, accounts for the lower hemoglobin levels among African-Americans? There are a number of sociological reasons that cause one subgroup to be subjected to more bodily insults than another, particularly in a society where racism is, unfortunately, all too common (Kent 1992). There are hematological data to support this contention. Later studies that performed serum ferritin measurements show that whereas African-Americans have lower hemoglobin levels, they concomitantly have higher serum ferritin levels, ranging from 271.7 ng/mL in black men to 111 ng/mL in black women; this contrasts to 93.63 in white men and 37.9 ng/mL in white women (Blumenfeld et al. 1988). Even though sedimentation rates (or ESR) were not elevated among a sample of African-Americans in one study, there was a higher incidence of chronic diseases as indicated by the number of serum ferritin levels above 100 ng/mL, which, in combination with low hemoglobin levels, defines the anemia of chronic disease/inflammation. Other investigations corroborate these findings. Of 78 African-Americans studied, 17.6 percent had elevated serum ferritin levels, leading the investigators to conclude that in addition to the anemia of chronic disease, there also were a number of individuals with iron overload who also suffered from occult inflammatory conditions (Haddy et al. 1986: 1084). The reason that more blacks than whites have the anemia of chronic disease is probably related to poverty. Many suffer from overcrowding, inadequate shelter, poor medical care, and the stress of minority status.

Black and white economic levels were matched in the studies mentioned. However, blacks within the same level of income as whites may suffer more stresses and infectious diseases because of their minority position and associated problems, such as prejudice and alcoholism, which were not taken into account in these studies. Other research indicates that whereas blacks have a lower hematocrit, it is only 0.7 percent lower than that of whites; R. Yip, S. Schwartz, and A. Deinard report the the "lower value in blacks may be accounted for by mild thalassemias, which are associated with lower hematocrit values. The use of the same diagnostic criteria for anemia among all

races will permit uniform detection of nutritional anemia" (Yip, Schwartz, and Deinard 1984: 824). The higher incidence of thalassemias associated with blacks in the United States is a geographical, not a racial, trait since whites indigenous to Mediterranean regions cursed in the past with endemic malaria also have a higher rate of thalassemias.

Various other minorities in North America are impoverished and suffer from prejudice, which is reflected in their hematology. The prevalence of anemia among Chinese-Canadians is similar to that of black Americans but dissimilar to that of white Americans (Chan-Yip and Gray-Donald 1987). Hispanic females between the ages of 20 and 44 had statistically higher levels of anemia than whites or blacks, although it is unfortunate that the etiology of the anemia was not determined (Looker et al. 1989). Thus, lower hemoglobin levels are not specifically a black-associated trait but do appear to be a minority-associated trait.

Native Americans are similarly affected with significantly higher incidences of anemia. Between 22 and 28 percent of Alaskan Native Americans were found to be anemic: Of these, 65 percent had iron deficiency and 35 percent had the anemia of chronic disease (Centers for Disease Control 1988). Parasites were not investigated in these studies, but earlier studies of this group indicate that they are common and a potential reason for the reported high frequency of iron deficiency (Rausch 1951; Schiller 1951). Pneumococcal disease also is endemic among Alaskan natives (Davidson et al. 1989). Poverty and poor nontraditional diets, combined with overcrowding in semisedentary and sedentary villages, contribute to this unfortunate situation. As discussed in the section "Elsewhere in the United States," skeletal material reveals that anemia was common in this population in the past as well, even though large quantities of heme iron in the form of meat were routinely consumed. Such anemia was the result of endemic health problems associated with winter village life and infection with parasites from contact with infected dogs, seals, polar bears, and other animals with which native Alaskans were in contact (Kent 1986).

By contrast, Native American children in Arizona were reported to have lower rates of anemia (Yip et al. 1987). This is difficult to assess because information was not provided as to which group of Native Americans was involved in the study. The frequency of anemia could be related to the dispersed settlement patterns of some groups, such as the Navajos, or to the season when the study was conducted, because many Native American children attend boarding schools during the winter but spend the summer at home. The frequency might also be attributed to the length of time children are breast-fed, which, at least in the past, was of longer duration than among Euroamerican children. In other studies,

Native Americans had the same relatively low rate of anemia as Euroamerican children (Yip, Schwartz, and Deinard 1984). Whatever the cause, the lower incidence of anemia among Arizona Native Americans in contrast to Alaskan Native Americans again demonstrates that anemia is not related to genetic or racial factors but is related to environmental and sociopolitical factors, such as poverty and its associated diseases, like alcoholism.

Anemia and Alcoholism, AIDS, Drugs

Anemia, primarily of chronic diseases, is correlated with a number of health problems currently affecting all countries to various degrees. Although alcohol enhances iron absorption and can cause overload and consequent pathologies (Rodriguez et al. 1986), alcoholics may also suffer from anemia. Between 13 to 62 percent of chronic alcoholics are anemic, primarily as a result of acute blood loss and illness, including liver disease (Savage and Lindenbaum 1986).

AIDS, or acquired immunodeficiency syndrome, is associated with anemia, but its interpretation is difficult. Anemia may be the result of the body's defense against the virus or its defense against the many secondary infections associated with AIDS. Anemia in AIDS patients can also be partly related to malnourishment from malabsorption associated with the condition. The latter is the least likely explanation, however, because the anemia is associated with normal or increased serum ferritin levels (Beutler 1988).

Secondary infections are more likely to cause anemia in this group. Many AIDS patients suffer from neoplasia of various types (especially Kaposi sarcoma and non-Hodgkin's lymphoma) and from a wide range of bacterial infections that are particularly virulent because of the host's weakened resistance (Brasitus and Strin 1990). As a result of their compromised immunological systems, AIDS patients are often afflicted with atypical mycobacterial infections as well (Ries, White, and Murdock 1990). Whatever the ultimate cause of the anemia, it appears that the hypoferrremia associated with AIDS is primarily the anemia of chronic disease and occurs as the body attempts to defend itself against very formidable infections.

Anemia and Performance

It has long been suggested that iron-deficiency anemia adversely affects performance, as measured by a battery of tests designed to evaluate activity, ability, and endurance (Dallman 1989). In fact, many studies claim to demonstrate a relationship between poor mental or physical performance and low hemoglobin level. However, a number of these studies indicate that changes noted were not statistically significant, did not include a control group, or were ambiguous in what prompted the improvements noted (Lozoff and Brittenham 1987; also see Lozoff, Jimenez, and Wolf 1991, who indicate the difficulties in determining cause and effect). The problem encountered here,

as throughout this discussion of anemia, is the separation of cause and consequence: physiological problem or defense.⁵

Furthermore, serious questions exist in almost every study that interprets behavioral, mental, or other functional limitations associated with dietary iron-deficiency anemia. For instance, no one denies that someone severely iron deficient as a result of blood loss will perform poorly compared to a healthy person. However, when anemia is reversed, other health problems are also corrected. Intake of calories, protein, and vitamins is improved in most of the studies designed to investigate performance and anemia. In the case of anemia of chronic disease, disease and/or parasite loads were reduced as the result of medication. Is it the increase in iron that is causing the improvement in skills, or is it the improvement in overall health and nutrition? Ingestion of oral iron can stimulate appetite, and there is often a weight gain associated with higher hemoglobin levels (Kent et al. 1990). It is this overall improvement in nutrition that has been suggested as the cause of the improved mental faculties (Wadsworth 1992).

Iron is necessary to catalyze various facets of the humoral and cell-mediated immune systems (Weinberg 1984). However, severely deficient individuals are not usually part of the performance test groups reported in the literature. Although iron-deficient individuals have a lower maximum work capability and endurance (Gardner et al. 1977; Lozoff and Brittenham 1987), we again are left with questions. How much of poor performance is related to iron deficiency and how much is related to poor calorie, vitamin, and protein intake and to parasitic or other disease?

Other research cannot be assessed. For example, one study compared anemic and nonanemic Chilean infants by conducting various mental and psychomotor tests that included talking and walking. Anemic infants performed more poorly than nonanemic ones (Walter et al. 1989). The difficulty in interpreting this finding is that the mental and psychomotor skills were calculated for each group according to their iron status and not according to gender. Female infants, in general, mature more rapidly than males; therefore, the gender composition of each group is vital for determining if the nonanemic group performed better than the anemic one simply because it was composed of more female infants, regardless of the presence or absence of anemia. After three months of iron supplementation, which reversed the anemia, "no significant improvement was detected between the scores at 12 and 15 months of age" (Walter et al. 1989: 12). This suggests either that the gender composition of the groups may have affected initial performance rates or that the anemia itself did not significantly affect the infants' ability on the tests.

Many studies, particularly earlier ones, also do not measure serum ferritin levels but simply define all

anemia as iron deficiency from poor diets and ignore improvements in health during the study (Soemantri, Pollitt, and Kim 1985; Chwang, Soemantri, and Pollitt 1988). In one study, anemic and nonanemic Thai fifth-grade students received either an iron supplement or a placebo concurrently with an anthelmintic drug to kill parasites (Pollitt et al. 1989). In addition, all other infections were treated on the first day of the study. Consequently, the end results are not comparable to the initial test results; that is, the cause of changes could be attributed to improved general health as much as to improved iron status.

As noted by Moussa Youdim (1989), test results often implicate factors other than iron, such as social and economic status, as the causes of differences in anemic and iron-replete students' performance before treatment. Even without taking this into account, studies indicate no difference in performance between treated and untreated children, and research "provides no support for an assumption of causality [between test scores and iron supplementation]" (Pollitt et al. 1989:695).

Iron-deficient infants in Costa Rica who were weaned early (average 4.9 months) or never breast-fed (16 percent) were given cow's milk, known to cause gastrointestinal bleeding in a high percentage of infants (Lozoff et al. 1987). These children performed less well on a series of mental and motor tests than did nonanemic infants (Lozoff et al. 1987). Later, after their anemia was corrected with iron supplements, their performances improved (Lozoff et al. 1987). It is suggested that rectifying blood losses, rather than diet, made the difference.

Nondietary causes of anemia are prevalent in many of the populations studied, as in New Guinea, where both malaria (and anemia of chronic disease) and thalassemia are common. A series of tests to measure attention span were given to New Guinea infants: malaria-positive and negative and iron-dextran supplementation with malaria and without (Heywood et al. 1989). The only significant difference recorded was that malaria-positive infants, regardless of iron supplementation, had longer fixation times than did infants with no evidence of malaria, regardless of iron supplementation (Heywood et al. 1989). As Horowitz (1989) points out, the precise meaning of longer fixation times is not known; differences may not reflect mental development but the opposite, or may simply be a reflection of age, maternal education, or general health status.

The results of another study of iron-deficient children in Indonesia was also ambiguous. After receiving iron supplementation, neither anemic nor iron-replete children significantly improved performance on some tests, but both groups improved on other tasks (Soewondo, Husaini, and Pollitt 1989). Factors not investigated, such as increases in caloric intake, changes in psychological state, or normal physiological maturation, may improve scores, regardless of anemia. If such

is the case, it creates problems with many of the studies that attempt to measure anemia and performance. An insightful commentary by Betsy Lozoff on this study concludes that "at this stage in the research on the behavioral effects of ID [iron-deficiency anemia], it seems reasonable to keep asking whether alterations in affect, motivation, or fatigue might underlie cognitive-test-score findings" (Lozoff 1989:675).

Anemia and Sports

An interesting correlation exists between marathon runners and subnormal serum ferritin levels. Although a high cutoff of less than 20 ng/mL was used to determine iron deficiency among 30 female high school long-distance runners, 45 percent were classified as anemic; among 10 other female runners, 50 to 60 percent were classified as anemic (Rowland et al. 1988; Manore et al. 1989). Several studies indicate that regardless of dietary iron intake, individuals who ran the most miles per week had the lowest ferritin levels; if injury that prevented further running occurred, iron indexes rose (Manore et al. 1989).

Lowered serum ferritin levels in athletes have been attributed to many factors: increased iron loss through heavy sweating; trauma; slightly increased destruction of red blood cells in stressed tissues, such as muscles and the soles of the feet; common use of analgesics, such as aspirin and aspirinlike drugs, which cause blood loss; and gastrointestinal blood loss (Robertson, Maughan, and Davidson 1987). A significant, though clinically unimportant, increase in fecal blood loss occurred in male marathon runners who had not taken any drugs prior to running. The 28 percent who had taken an analgesic known to promote bleeding had blood losses that could eventually result in anemia (Robertson et al. 1987). Most physicians consider the blood loss and its cause inconsequential in healthy athletes and do not recommend routine hematological monitoring or iron supplementation (Wardrop 1987).

Conclusions and Direction of Future Studies

Despite many studies supporting the view that iron fortification and supplementation might be harmful for infants, pregnant women, and others (for example, Hibbard 1988), some physicians continue to advocate indiscriminate iron fortification (Arthur and Isbister 1987; Taylor and Bergman 1989). However, a growing number of studies concerning the anemia of chronic disease conclude that "[o]ur data . . . do not support the routine prescription of iron . . . in patients on CAPD [continuous, ambulatory peritoneal dialysis]" (Salahudeen et al. 1988). In fact, even though the use of iron chelators that bind iron to reduce chronic diseases is still experimental, initial studies look promising (Vreugdenhil et al. 1989). That is, less iron rather than more may reduce morbidity.

Treating all anemia as dietary iron-deficiency anemia can have potentially deleterious effects on popu-

lations most needing health-care assistance. Iron-deficiency anemia occurs in areas where war, export or cash cropping, and other extreme situations deny people their basic requirements of calories, vitamins, and nutrients to sustain life. Dietary improvements are sorely needed in those situations. Anemia from blood loss primarily from parasites, and also from the premature introduction of cow's milk to infants, is an acute health-maintenance problem. Rectification through medication and improved sanitation and education is needed. Anemia of chronic disease is a positive defense against infections and inflammation and, as such, should not be interfered with; however, the underlying diseases that cause the anemia need to be eradicated.

Iron supplementation is a relatively easy, inexpensive counter to dietary anemia. Perhaps that partially explains why so many people cling to the idea that iron taken orally or through injections will reduce morbidity. However, the complexity of iron and its relationship to anemia are emphasized here to show that simple solutions cannot solve complex problems.

Susan Kent

This chapter is dedicated to Dr. Eugene Weinberg and his work. I greatly value the comments and suggestions made by Drs. Steven Kent and Weinberg. I thank Dr. Kenneth Kiple for giving me the opportunity to write this paper and for editorial advice. Marian Blue also provided editorial suggestions.

Notes

1. One study (Baltimore, Shedd, and Pearson 1982) contends that studies in vivo do not support the conclusions from studies in vitro of excess iron predisposing individuals to infections. One explanation for their finding that increased iron does not increase *E. coli* growth is that a few bacteria, including *E. coli* possessing K₁ antigen, are virulent even without iron supplementation (Payne and Finkelstein 1978). This group of pathogens "appeared to produce greater quantities of compounds (siderophores) which stimulated microbial growth in low-iron media than did the nondisseminating pathogens" (Payne and Finkelstein 1978: 1428). Unfortunately, the Baltimore, Shedd, and Pearson (1982) study does not specify whether the *E. coli* they investigated contained the K₁ antigen, but it is possible that other mutants of a bacterium may be able to proliferate in low-iron media. The vast majority of other studies contradict these authors.
2. Although at one time it was thought that hookworm was not indigenous to the Americas, hookworm was found in a 1,100-year-old Peruvian mummy and in a 3,500-year-old Brazilian mummy (Allison 1974; Ferreira, de Araujo, and Confalonieri 1980, 1983).
3. The same hematological surveys detected iron overload in 1.3 percent of 157 healthy urban adults, a figure similar to the percentage of anemic individuals for males (Haddy et al., 1986). However, in a larger study that involved 11,065 subjects, iron overload as measured by elevated transferrin saturation levels was detected in only 0.8 percent of men and 0.3 percent of women (Edwards et al. 1988). The problem with interpreting this study is that the subject population was drawn from blood donors, a practice that tends to lower transferrin saturation levels. The actual prevalence of iron overload appears to be roughly similar to anemia in males under the age of 45.
4. There are a few specific genes associated with particular populations within black or other racial categories but not genes that only blacks share to the exclusion of all other groups.
5. A similar problem of distinguishing cause and effect can be seen with the immunological system and anemia. Some investigators claim their studies illustrate that anemia causes an impairment in the immunological system, which affects the body's ability to ward off disease (e.g., Vyas and Chandra 1984). Others claim their studies illustrate that anemia improves immunity from disease (Murray and Murray 1977). For instance, despite in vitro studies that show experimentally that there is a decrease in bactericidal capabilities in neutrophils, in vivo studies of human infections show that "no noticeable increase was observed in respiratory, gastrointestinal, or in morbidity, either in the few days before the initiation of iron therapy, or in the subsequent 15 days of close clinical and laboratory follow-up, confirming previous observations that iron-deficiency anemia, even when severe, does not appear to compromise immune mechanisms to the extent of allowing ominous clinical manifestations" (Walter et al. 1986).

Bibliography

- Aaby, Peter. 1988. Observing the unexpected: Nutrition and child mortality in Guinea-Bissau. In *Micro-approaches to demographic research*, ed. John Caldwell, Allen Hill, and Valerie Hull, 278–96. London.
- Adelekan, Delana, and David Thurnham. 1990. Plasma ferritin concentrations in anemic children: Relative importance of malaria, riboflavin deficiency, and other infections. *American Journal of Clinical Nutrition* 51: 453–6.
- Allison, Marvin J. 1984. Paleopathology in Peruvian and Chilean populations. In *Paleopathology at the origins of agriculture*, ed. Mark Cohen and George Armelagos, 515–29. Orlando, Fla.
- Allison, Marvin J., Alejandro Pezzia, Ichiro Hasegawa, et al. 1974. A case of hookworm infestation in a pre-columbian American. *American Journal of Physical Anthropology* 41: 103–6.
- Anderson, James. 1968. Late Paleolithic skeletal remains from Nubia. In *The prehistory of Nubia*, ed. Fred Wendorf, 996–1040. Dallas, Tex.
- Angel, J. Lawrence. 1967. Porotic hyperostosis or osteoporosis. In *Diseases in antiquity*, ed. Don Brothwell and A. T. Sandison, 378–89. Springfield, Ill.
1984. Health as a crucial factor in the changes from hunting to developed farming in the eastern Mediterranean. In *Paleopathology at the origins of agriculture*, ed. Mark Cohen and George Armelagos, 51–74. Orlando, Fla.
- Armelagos, George, Dennis van Gerven, Debra Martin, and Rebecca Huss-Ashmore. 1984. Effects of nutritional change on the skeletal biology of northeast African (Sudanese Nubian) populations. In *From hunters to farmers: The causes and consequences of food production in Africa*, ed. J. Desmond Clark and Steven Brandt, 132–46. Berkeley, Calif.
- Arthur, C. K., and J. P. Isbister. 1987. Iron deficiency: Misunderstood, misdiagnosed and mistreated. *Drugs* 33: 171–82.
- Aufderheide, Arthur. 1989. Chemical analysis of skeletal remains. In *Reconstruction of life from the skeleton*, ed. Mehmet Iscan and Kenneth Kennedy, 237–60. New York.

- Aung-Thau-Batu, M. B., U. Hla-Pe, M. Thein-Than, and M. B. Khin-Kyi-Nyunt. 1972. Iron deficiency in Burmese population groups. *American Journal of Clinical Nutrition* 25: 210-18.
- Baldwin, J. Gilbert. 1989. True anemia: Incidence and significance in the elderly. *Geriatrics* 44: 33-6.
- Baltimore, R., D. Shedd, and H. Pearson. 1982. Effect of iron saturation on the bacteriostasis of human serum: In vivo does not correlate with in vitro saturation. *Journal of Pediatrics* 101: 519-23.
- Barry, D. M., and A. W. Reeve. 1977. Increased incidence of gram-negative neonatal sepsis with intramuscular iron administration. *Pediatrics* 60: 908-12.
- Bates, C. J., H. J. Powers, and D. I. Thurnham. 1989. Vitamins, iron, and physical work. *Lancet* 2: 313-14.
- Becroft, D. M., M. R. Dix, and K. Farmer. 1977. Intramuscular iron-dextran and susceptibility of neonates to bacterial infections. *Archives of Diseases of Childhood* 52: 778-81.
- Beganovic, Sead. 1987. The erythrocyte sedimentation rate. *Annals of Internal Medicine* 107: 425-6.
- Bentley, D. P. 1985. Iron metabolism and anaemia in pregnancy. *Clinics in Haematology* 14: 613-28.
- Beutler, Ernest. 1988. The common anemias. *Journal of the American Medical Association* 259: 2433-7.
- Bindra, Gursh, and Rosalind Gibson. 1986. Iron status of predominantly lacto-ovo vegetarian East Indian immigrants to Canada: A model approach. *American Journal of Clinical Nutrition* 44: 643-52.
- Blumenfeld, Nadia, Mary Fabry, Benjamin Thysen, and Ronald Nagel. 1988. Red cell density is sex and race dependent in the adult. *Journal of Laboratory and Clinical Medicine* 112: 333-8.
- Brasitus, Thomas A., and Michael D. Sitrin. 1990. Intestinal malabsorption syndromes. *Annual Review of Medicine* 41: 339-47.
- Brennan, Ruth, Mary Kohrs, James Nordstrom, et al. 1983. Nutrient intake of low-income pregnant women: Laboratory analysis of foods consumed. *Journal of the American Dietetic Association* 83: 546-50.
- Bumsted, M. Pamela. 1980. VT-CH-94: Vermont's earliest known agricultural experiment station. *Man in the Northeast* 19: 73-82.
1985. Past human behavior from bone chemical analysis - Respects and prospects. *Journal of Human Evolution* 14: 539-51.
- Burns, Edward, S. N. Goldberg, Christine Lawrence, and Barry Weinz. 1990. Clinical utility of serum tests for iron deficiency in hospitalized patients. *American Journal of Clinical Pathology* 93: 240-5.
- Cash, Joseph, and David Sears. 1989. The anemia of chronic disease: Spectrum of associated diseases in a series of unselected hospitalized patients. *American Journal of Medicine* 87: 638-44.
- Centers for Disease Control and Prevention U.S. 1986. Declining anemia prevalence among children enrolled in public nutrition and health programs: Selected states, 1975-1985. *Morbidity and Mortality Weekly Report* 35: 565-7.
1988. High prevalence of iron deficiency anemia among Alaskan native Children. *Morbidity and Mortality Weekly Report* 37: 200-2.
- Chan-Yip, Alice, and Katherine Gray-Donald. 1987. Prevalence of iron deficiency among Chinese children aged 6 to 36 months in Montreal. *Canadian Medical Association Journal* 136: 373-8.
- Charache, Samuel, et al. 1987. Noninvasive assessment of tissue iron stores. *American Journal of Clinical Pathology* 88: 333-7.
- Charoenlarp, P., S. Dhanamitta, R. Kaewvichit, et al. 1988. A WHO collaborative study on iron supplementation in Burma and in Thailand. *American Journal of Clinical Nutrition* 47: 280-97.
- Chernela, Janet, and Vernon Thatcher. 1989. Comparison of parasite burdens in two native Amazonian populations. *Medical Anthropology* 10: 279-85.
- Chwang, Leh-chii, A. G. Soemantri, and Ernesto Pollitt. 1988. Iron supplementation and physical growth of rural Indonesian children. *American Journal of Clinical Nutrition* 47: 496-501.
- Clydesdale, Fergus. 1983. Physicochemical determinants of iron bioavailability. *Food Technology* 133-44.
- Cohen, Mark and George Armelagos, eds. 1984. *Paleopathology at the origins of agriculture*. Orlando, Fla.
- Cook, Della. 1984. Subsistence and health in the lower Illinois Valley: Osteological evidence. In *Paleopathology at the origins of agriculture*, ed. Mark Cohen and George Armelagos, 235-69. Orlando, Fla.
- Cook, James. 1990. Adaption in iron metabolism. *American Journal of Clinical Nutrition* 51: 301-8.
- Cook, James, et al. 1971. Nutritional deficiency and anemia in Latin America: A collaborative study. *Blood* 38: 591-603.
- Cook, James, and Sean Lynch. 1986. The liabilities of iron deficiency. *Blood* 68: 803-9.
- Cook, James, and Barry Skikne. 1982. Serum ferritin: A possible model for the assessment of nutrient stores. *American Journal of Clinical Nutrition* 35: 1180-5.
- Cook, James, Barry Skikne, Sean Lynch, and Molly Reusser. 1986. Estimates of iron sufficiency in the U.S. population. *Blood* 68: 726-31.
- Corwin, Edwards, Linda Griffin, David Goldgar, et al. 1988. Prevalence of hemochromatosis among 11,065 presumably healthy blood donors. *New England Journal of Medicine* 318: 1355-62.
- Cule, John, and I. Lynn Evans. 1968. Porotic hyperostosis and the Gelligaer skull. *Journal of Clinical Pathology* 21: 753-8.
- Cummings, Linda. 1989. Coprolites from medieval Christian Nubia: An interpretation of diet and nutritional stress. Ph.D. dissertation, Department of Anthropology, University of Colorado, Boulder.
- Dallman, Peter. 1987. Has routine screening of infants for anemia become obsolete in the United States? *Pediatrics* 80: 439-41.
1989. Iron deficiency: Does it matter? *Journal of Internal Medicine* 226: 367-72.
- Dallman, Peter, and Jerry Reeves. 1984. Laboratory diagnosis of iron deficiency. In *Iron nutrition in infancy and childhood*, ed. A. Stekel, 11-43. New York.
- Dallman, Peter, and Ray Yip. 1989. Changing characteristics of childhood anemia. *The Journal of Pediatrics* 114: 161-4.
- Dallman, Peter, Ray Yip, and Clifford Johnson. 1984. Prevalence and causes of anemia in the United States, 1976 to 1980. *American Journal of Clinical Nutrition* 39: 437-45.
- Daly, Mel. 1989. Anemia in the elderly. *American Family Physician* 39: 129-36.
- Daly, Mel, and Jeffery Sobal. 1989. Anemia in the elderly: A survey of physicians' approaches to diagnosis and workup. *Journal of Family Practice* 28: 524-8.
- Davidson, Michael, Cynthia Schraer, Alan Parkinson, et al. 1989. Invasive pneumococcal disease in an Alaskan native population, 1980 through 1986. *Journal of the American Medical Association* 261: 715-18.
- Derman, D. P., et al. 1978. Serum ferritin as an index of iron nutrition in rural and urban South African children. *British Journal of Nutrition* 39: 383-9.

- Dunn, Frederick. 1968. Epidemiological factors: Health and disease in hunter-gatherers. In *Man the hunter*, ed. Richard Lee and Irven DeVore, 221–8. Chicago.
- Edwards, Corwin, et al. 1988. Prevalence of hemochromatosis among 11,065 presumably healthy blood donors. *New England Journal of Medicine* 318: 1355–62.
- El-Najjar, Mahmoud Y. 1976. Maize, malaria and the anemias in the pre-Columbian New World. *Yearbook of Physical Anthropology* 20: 329–37.
- Expert Scientific Working Group. 1985. Summary of a report on assessment of the iron nutritional status of the United States population. *American Journal of Clinical Nutrition* 42: 1318–30.
- Fairbanks, Virgil, and Ernest Beutler. 1988. Iron. In *Modern nutrition in health and disease*, ed. Maurice Shils and Vernon Young, 193–226. Philadelphia, Pa.
- Farley, Patrick, and Jaime Foland. 1990. Iron deficiency anemia: How to diagnose and correct. *Postgraduate Medicine* 87: 89–101.
- Fernandes-Costa, F. J., John Marshall, Clair Ritchie, et al. 1984. Transition from a hunter-gatherer to a settled lifestyle in the 'Kung San: Effect on iron, folate, and vitamin B12 nutrition. *The American Journal of Clinical Nutrition* 40: 1295–1303.
- Ferreira, L. F., A. J. de Araujo, and U. Confalonieri. 1980. The finding of eggs and larvae of parasitic helminths in archaeological material from Unai, Minas Gerais, Brazil. *Transcripts of the Royal Society for Tropical and Medical Hygiene* 74: 780–800.
1983. The finding of helminth eggs in a Brazilian mummy. *Transcripts of the Royal Society for Tropical and Medical Hygiene* 77: 65–7.
- Finch, C.A. 1989. Introduction: Knights of the oval table. *Journal of Internal Medicine* 226:345–8.
- Florentino, Rodolfo, and Romualda Guirriec. 1984. Prevalence of nutritional anemia in infancy and childhood with emphasis on developing countries. In *Iron nutrition in infancy and childhood*, ed. A. Stekel, 61–74. New York.
- Fomon, S. J., E. E. Ziegler, S. E. Nelson, and B. B. Edwards. 1981. Cow milk feeding in infancy: Gastrointestinal blood loss and iron nutritional status. *Journal of Pediatrics* 98: 540–5.
- Foucard, T. 1985. Development of food allergies with special reference to cow's milk allergy. *Pediatrics* 75: 177–81.
- Franke, Richard. 1987. The effects of colonialism and neocolonialism on the gastronomic patterns of the Third World. In *The evolution of food*, ed. Marvin Harris and Eric Ross, 455–79. Philadelphia, Pa.
- Garby, L., L. Irnell, and I. Werner. 1969. Iron deficiency in women of fertile age in a Swedish communities, III: Estimation of prevalence based on response to iron supplementation. *Acta Medica Scandinavica* 185: 113–30.
- Gardner, Gerald, V. Reggie Edgerton, Brian Senewiratne, et al. 1977. Physical work capacity and metabolic stress in subjects with iron deficiency anemia. *American Journal of Clinical Nutrition* 30: 910–17.
- Garn, Stanley, and Diane Clark. 1976. Problems in the nutritional assessment of black individuals. *American Journal of Public Health* 66: 262–7.
- Garn, Stanley, Helen Shaw, and Kinne McCabe. 1976. Black-white hemoglobin differences during pregnancy. *Ecology of Food and Nutrition* 5: 99–100.
- Garn, Stanley, Nathan Smith, and Diane Clark. 1975. Lifelong differences in hemoglobin levels between blacks and whites. *Journal of the National Medical Association* 67: 91–6.
- Gilfillan, S. C. 1965. Lead poisoning and the fall of Rome. *Journal of Occupational Medicine* 7: 53–60.
- Goepel, E., H. U. Ulmer, and R. D. Neth. 1988. Premature labor contractions and the value of serum ferritin during pregnancy. *Gynecologic and Obstetric Investigation* 26: 265–73.
- Goldberg, A., and A. Reshef. 1972. Vitamin A and iron in infants' diets in Israel. *Journal of the American Dietetic Association* 60: 127–30.
- Goldstein, Marcus. 1957. Skeletal pathology of early Indians in Texas. *American Journal of Physical Anthropology* 15: 299–307.
- Goodman, Alan, John Lallo, George Armelagos, and Jeremy Rose. 1984. Health changes at Dickson Mounds, Illinois (A.D. 950–1300). In *Paleopathology at the origins of agriculture*, ed. Mark Cohen and George Armelagos, 271–305. Orlando, Fla.
- Guyatt, Gordon, Christopher Patterson, Mahmoud Ali, et al. 1990. Diagnosis of iron-deficiency anemia in the elderly. *American Journal of Medicine* 88: 205–9.
- Haddy, Theresa, Oswaldo Castro, Sohail Rana, and Roland Scott. 1986. Iron status and liver function in healthy adults: A multiracial pilot study. *Southern Medical Journal* 79: 1082–5.
- Hallberg, Leif, Mats Brune, and Lena Rossander. 1989. Iron absorption in man: Ascorbic acid and dose-dependent inhibition by phytate. *American Journal of Clinical Nutrition* 49: 140–4.
- Hemminki, Elina. 1978. Routine administration of iron and vitamins during pregnancy: Review of controlled clinical trials. *British Journal of Obstetrics and Gynecology* 85: 404–10.
- Hengen, O. P. 1971. Cribra orbitalia: Pathogenesis and probable etiology. *Homo* 22: 57.
- Herbert, Victor. 1988. Recommended dietary intakes (RDI) of iron in humans. *American Journal of Clinical Nutrition* 45: 679–86.
- Herold, P. M., and J. E. Kinsella. 1986. Fish oil consumption and decreased risk of cardiovascular disease. *American Journal of Clinical Nutrition* 43: 566–98.
- Heywood, Alison, Stephen Oppenheimer, Peter Heywood, and Damien Jolley. 1989. Behavioral effects of iron supplementation in infants in Madang, Papua New Guinea. *American Journal of Clinical Nutrition* 50: 630–40.
- Hibbard, Bryan. 1988. Iron and folate supplements during pregnancy: Supplementation is valuable only in selected patients. *British Medical Journal* 297: 1324–7.
- Hoffbrand, A. V. 1981. Iron. In *Postgraduate hematology*, ed. A. V. Hoffbrand and S. M. Lewis, 35–42. New York.
- Hooton, Ernest. 1940. Skeletons from the Cenote of Sacrifice at Chichén Itzá. In *The Maya and their neighbors*, ed. Clarence Hay, 272–80. New York.
- Hornabrook, R. W. 1977. Human adaptability in Papua New Guinea. In *Population structure and human variation*, ed. G. A. Harrison, 285–309. Cambridge.
- Horowitz, Frances. 1989. Comments on Heywood et al. *American Journal of Clinical Nutrition* 50: 630–40.
- Houwelingen, R. V., A. Nordoy, E. van der Beek, et al. 1987. Effect of a moderate fish intake on blood pressure, bleeding time, hematology, and clinical chemistry in healthy males. *American Journal of Clinical Nutrition* 46: 424–36.
- Hummert, J.R. 1983. Cortical bone growth and dietary stress among subadults from Nubia's Batn el Hajar. *American Journal of Physical Anthropology* 62: 167–76.
- Jackson, Robert, and Linda Jackson. 1987. Biological and

- behavioral contributors to anemia during pregnancy in Liberia, West Africa. *Human Biology* 59: 585-97.
- Janssens, Paul. 1970. *Palaeopathology: Diseases and injuries of prehistoric man*. London.
- Jansson, L. T., S. Kling, and P. R. Dallman. 1986. Anemia in children with acute infections seen in a primary care pediatric outpatient clinic. *Pediatric Infectious Diseases* 5: 424-7.
- Keegan, William. 1989. Stable isotope analysis of prehistoric diet. In *Reconstruction of life from the skeleton*, ed. Mehmet Iscan and Kenneth Kennedy, 223-36. New York.
- Kennedy, Kenneth. 1984. Growth, nutrition and pathology in changing paleodemographic settings in South Asia. In *Paleopathology at the origins of agriculture*, ed. Mark Cohen and George Armelagos, 169-92. Orlando, Fla.
- Kent, Susan. 1986. The influence of sedentism and aggregation on porotic hyperostosis and anaemia: A case study. *Man* 21: 605-3.
1992. Anemia through the ages: Changing perspectives and their implications. In *Diet, demography, and disease: Changing views of anemia*, ed. Patricia Stuart-Macadam and Susan Kent, 1-32. New York.
1992. Excavations at a small Mesa Verde Pueblo II Anasazi site in southwestern Colorado. *The Kiva* 57.
- Kent, Susan, and David Dunn. 1991. A hematological study of a recently sedentary Kalahari village: Implications for the past and the future. [Manuscript submitted for publication].
1993. Etiology of hypoferrremia in a recently sedentary Kalahari village. *American Journal of Tropical Medicine and Hygiene* 48: 554-67.
1996. Anemia and the transition of nomadic hunter-gathers to a sedentary life-style: Follow-up study of a Kalahari community. *American Journal of Physical Anthropology* 99: 455-72.
- Kent, Susan, and Richard Lee. 1992. A hematological study of !Kung Kalahari foragers: An eighteen year comparison. In *Diet, demography, and disease: Changing views of anemia*, ed. Patricia Stuart-Macadam and Susan Kent, 173-200. New York.
- Kent, Susan, Eugene Weinberg, and Patricia Stuart-Macadam. 1990. Dietary and Prophylactic iron supplements: Helpful or harmful? *Human Nature* 1: 55-81.
1994. The etiology of the anemia of chronic disease and infection. *Journal of Clinical Epidemiology* 47: 23-33.
- Klebanoff, M., et al. 1989. Facts and artifacts about anemia and preterm delivery. *Journal of American Medical Association* 262: 511-15.
- Klepinger, Linda. 1984. Nutritional assessment from bone. *Annual Review of Anthropology* 13: 75-96.
- Kluger, M. J., and J. J. Bullen. 1987. Clinical and physiological aspects. In *Iron and Infection*, ed. J. J. Bullen and E. Griffiths, 243-82. London.
- Larrick, James, James Yost, Jon Kaplan, et al. 1979. Patterns of health and disease among the Waorami Indians of eastern Ecuador. *Medical Anthropology* 19: 147-89.
- Larson, Clark. 1984. Health and disease in prehistoric Georgia: The transition to agriculture. In *Paleopathology at the origins of agriculture*, ed. Mark Cohen and George Armelagos, 367-92. Orlando, Fla.
- Lawrence, Dale, James Neel, Stanley Abadie, et al. 1980. Epidemiological studies among Amerindian populations of Amazonia: Intestinal parasitosis in newly contacted and acculturating villages. *American Journal of Tropical Medicine and Hygiene* 29: 530-7.
- Lepowsky, M. 1985. Food taboos, malaria and dietary change: Infant feeding and cultural adaption on a Papua New Guinea Island. *Ecology of Food Nutrition* 16: 117.
- Lieberman, Ellice, Kenneth Ryan, Richard Monson, and Stephen Schoenbaum. 1988. Association of maternal hematocrit with premature labor. *American Journal of Obstetrics and Gynecology* 159: 107-14.
- Looker, Anne, Clifford Johnson, Margaret McDowell, and Elizabeth Yetley. 1989. Iron status: Prevalence of impairment in three Hispanic groups in the United States. *American Journal of Clinical Nutrition* 49: 553-8.
- Lozoff, Betsy. 1989. Comment on Soewondo, Husaini, and Pollitt. *American Journal of Clinical Nutrition* 50: 673-4.
- Lozoff, Betsy, and Gary Brittenham. 1987. Behavioral alterations in iron deficiency. *Hematology/Oncology Clinics of North America* 1: 449-64.
- Lozoff, Betsy, Gary Brittenham, Abraham Wolf, et al. 1987. Iron deficiency anemia and iron therapy effects on infant developmental test performance. *Pediatrics* 79: 981-95.
- Lozoff, B., E. Jimenez, and A. W. Wolf. 1991. Long-term developmental outcome of infants with iron deficiency. *New England Journal of Medicine* 325: 687-94.
- Macfarlane, Bruce, et al. 1988. Inhibitory effect of nuts on iron absorption. *American Journal of Clinical Nutrition* 47: 270-4.
- Malville, Nancy. 1987. Iron deficiency anemia among village women of the Middle Hills of Nepal. Ph.D. dissertation, University of Colorado, Boulder.
- Manore, Melinda M., Phyllis D. Bensfielder, Christine L. Wells, et al. 1989. Nutrient intakes and iron status in female long-distance runners during training. *Journal of the American Dietetic Association* 89: 257-9.
- Marshall, John, and Claire Ritchie. 1984. Where are the Ju/wasi of Nyae Nyae? Changes in a Bushman Society: 1958-1981. *University of Cape Town Centre for African Studies* 9: 1-187.
- Masawe, A. E., J. M. Muindi, and G. B. Swai. 1974. Infections in iron deficiency and other types of anaemia in the tropics. *Lancet* 2: 314-17.
- Metz, J., D. Hart, and H. C. Harpending. 1971. Iron, folate, and vitamin B₁₂ nutrition in a hunter-gatherer people: A study of the !Kung Bushmen. *The American Journal of Clinical Nutrition* 24: 229-42.
- Mills, A. F. 1990. Surveillance for anaemia: Risk factors in patterns of milk intake. *Archives of Disease in Childhood* 65: 428-31.
- Monsen, Elaine. 1988. Iron nutrition and absorption: Dietary factors which impact iron bioavailability. *Journal of the American Dietetic Association* 88: 786-90.
- Moseley, John. 1965. The paleopathologic riddle of "symmetrical osteoporosis." *American Journal of Roentgenology, Radium Therapy and Nuclear Medicine* 95: 135-42.
- Muñoz, Leda, Bo Lönnerdal, Carl Keen, and Kathryn Dewey. 1988. Coffee consumption as a factor in iron deficiency anemia among pregnant women and their infants in Costa Rica. *American Journal of Clinical Nutrition* 48: 645-51.
- Murray, M. J., and A. Murray. 1977. Starvation suppression and refeeding activation of infection: An ecological necessity? *Lancet* 1: 123-5.
- Murray, M. J., Anne Murray, and C. J. Murray. 1980a. The salutary effect of milk on amoebiasis and its reversal by iron. *British Medical Journal* 280: 1351-2.
- 1980b. An ecological interdependence of diet and disease? *American Journal of Clinical Nutrition* 33: 697-701.
- Murray, M. J., A. Murray, N. Murray, and M. Murray. 1976. Somali food shelters in the Ogaden famine and their impact on health. *Lancet* 1: 1283-5.
- Nathan, Hilel. 1966. Cribra orbitalia: A bone condition of the orbit of unknown nature: Anatomical study with etio-

- logical considerations. *Israel Journal of Medical Science* 2: 171-91.
- Nelson, Brue, et al. 1986. Effects of digenesis on strontium, carbon, nitrogen and oxygen concentration and isotopic composition of bone. *Geochimica et Cosmochimica Acta* 50: 1941-9.
- Olivares, Manuel, et al. 1989. Anemia of a mild viral infection: The measles vaccine as a model. *Pediatrics* 84: 851-5.
- O'Neil-Cutting, Mary, and William Crosby. 1987. Blocking of iron absorption by a preliminary oral dose of iron. *Archives of Internal Medicine* 147: 489-91.
- Oppenheimer, S. J., F. D. Gibson, S.B. MacFarlane, et al. 1986a. Iron supplementation increases prevalence and effects of malaria: Report on clinical studies in Papua New Guinea. *Transactions of the Royal Society of Tropical Medicine and Hygiene* 80: 603-12.
- Oppenheimer, S. J., S. B. MacFarlane, J.B. Moody, et al. 1986b. Effect of iron prophylaxis on morbidity due to infectious disease: Report on clinical studies in Papua New Guinea. *Transactions of the Royal Society of Tropical Medicine and Hygiene* 80: 596-602.
- Orlove, Benjamin. 1987. Stability and change in highland Andean dietary patterns. In *The evolution of food*, ed. Marvin Harris and Eric Ross, 481-515. Philadelphia, Pa.
- Pampiglione, Silvio, and Maria Ricciardi. 1986. Parasitological surveys of Pygmy groups. In *African Pygmies*, ed. Luigi Cavalli-Sforza, 153-165. Orlando, Fla.
- Payne, S., and R. Finkelstein. 1978. The critical role of iron in host-bacterial interactions. *Journal of Clinical Investigations* 61: 1428-40.
- Perzigian, Anthony, Patricia Tench, and Donna Braun. 1984. Prehistoric health in the Ohio River Valley. In *Paleopathology at the origins of agriculture*, ed. Mark Cohen and George Armelagos, 347-66. Orlando, Fla.
- Pochedly, Carl, and Steven May. 1987. Iron deficiency anemia in children. *American Family Physician* 35: 195-200.
- Pollitt, Ernesto, Phongjan Hathirat, Nittaya Kotchabhakdi, et al. 1989. Iron deficiency and educational achievement in Thailand. *American Journal of Clinical Nutrition* 50: 687-97.
- Powell, James, Stephen Weens and Nanette Wenger. 1965. The skull roentgenogram in iron deficiency anemia and in secondary polycythemia. *American Journal of Roentgenology, Radium Therapy, and Nuclear Medicine* 95: 143-7.
- Powell, Mary. 1988. *Status and health in prehistory: A case study of the Moundville chiefdom*. Washington, D.C.
- Rausch, Robert. 1951. Biotic interrelationships of helminth parasitism in Alaska. *Bulletin of the National Research Council* no. 122: 113-14.
- Reeves, J. D., R. Yip, V. Kiley, et al. 1984. Iron deficiency in infants: The influence of antecedent infection. *Journal of Pediatrics* 105: 874-9.
- Reinhard, Karl. 1992. Diet, demography, and disease: The coprolite evidence. In *Diet, demography, and disease: Changing perspectives of anemia*. ed. P. Stuart-Macadam and S. Kent, 219-58. New York.
- Ries, Kristen, George White, and Richard Murdock. 1990. Atypical mycobacterial infection caused by *Mycobacterium marinum*. *New England Journal of Medicine* 322: 633-4.
- Ritchey, A. Kim. 1987. Iron deficiency in children: Update on an old problem. *Postgraduate Medicine* 32: 59-69.
- Robbins, Louise. 1986. Biosocial maladaptation among prehistoric maize agriculturalists: The role of trace elements. In *Skeletal analysis in southeastern archaeology*, ed. Janet Levy, 97-102. North Carolina Archaeological Council Publication 24. Chapel Hill.
- Robertson, J. D., R. J. Maughan, and R. J. Davidson. 1987. Faecal blood loss in response to exercise. *British Medical Journal* 295: 303-4.
- Rodolfo, Florentino, and Romualda Guirriec. 1984. Prevalence of nutritional anemia in infancy and childhood with emphasis on developing countries. In *Iron nutrition in infancy and childhood*, ed. A. Stekel, 61-74. New York.
- Rodriguez, M. C., M.S. Henriquez, A.F. Turon, et al. 1986. Trace elements in chronic alcoholism. *Trace Elements in Medicine* 3: 164-7.
- Ross, Eric. 1987. Overview of trends in dietary variation from hunter-gatherer to modern capitalist society. In *The evolution of food*, ed. Marvin Harris and Eric Ross, 7-55. Philadelphia, Pa.
- Rowland, Thomas, Molly Deisroth, Gerald Green, and John Kelleher. 1988. The effect of iron therapy on the exercise capacity of nonanemic iron-deficient adolescent runners. *American Journal of Childhood* 142: 165-9.
- Saarinen, Ulla, and Martti Siimes. 1977. Developmental changes in serum iron, total iron-binding capacity, and transferrin saturation in infancy. *Journal of Pediatrics* 91: 875-7.
- Salahudeen, A. K., S.R. Varma, T. Karim, et al. 1988. Anaemia, ferritin, and vitamins in continuous ambulatory peritoneal dialysis. *Lancet* 1: 1049.
- Saul, Frank. 1973. Disease in the Mayan era: The pre-Columbian evidence. In *The classic Maya collapse*. ed. Patrick Culbert, 301-24. Albuquerque, NM.
- Savage, David, and John Lindenbaum. 1986. Anemia in alcoholics. *Medicine* 65: 322-38.
- Schiller, Everett. 1951. Observations on arctic parasitology with special reference to the status of Echinococcosis and Diphyllorhynchiasis in Alaska. *Bulletin of the National Research Council* 122: 91-2.
- Schwaner, Terry, and Carl Dixon. 1974. Helminthiasis as a measure of cultural change in the Amazon Basin. *Biotropica* 6: 32-7.
- Shahidi, Nasollah, and Louis Diamond. 1960. Skull changes in infants with chronic iron-deficiency anemia. *New England Journal of Medicine* 262: 137-9.
- Sillen, Andrew, Judith Sealy, and Nikolaas van der Merwe. 1989. Chemistry and paleodietary research: No more easy answers. *American Antiquity* 54: 504-12.
- Smith, Patricia, Ofer Bar-Yosef, and Andrew Sillen. 1984. Archaeological and skeletal evidence for dietary change during the Late Pleistocene/Early Holocene in the Levant. In *Paleopathology at the origins of agriculture*, ed. Mark Cohen and George Armelagos, 101-36. Orlando, Fla.
- Soemantri, A. G., Ernesto Pollitt, and Insum Kim. 1985. Iron deficiency anemia and educational achievement. *American Journal of Clinical Nutrition* 42: 1221-8.
- Soewondo, Soesmalijah, M. Husaini, and Ernesto Pollitt. 1989. Effects of iron deficiency on attention and learning processes in preschool children: Bandung, Indonesia. *American Journal of Clinical Nutrition* 50: 667-74.
- Stander, Paul. 1989. Anemia in the elderly: Symptoms, causes, and therapies. *Postgraduate Medicine* 85: 85-96.
- Stonich, Susan. 1988. Integrated socioeconomic, remote sensing, and information management procedures for rural development and agricultural policy design. Paper presented at the 1988 Annual Meeting of the Society for Applied Anthropology, Tampa, Fla.
1991. The political economy of environmental destruction: Food security in Honduras. In *Harvest of want: Food*

- security in Mexico and Central America. ed. Scott Whiteford and Anne Ferguson, 35–55. Boulder, Colo.
- Stuart-Macadam, Patricia. 1982. A correlative study of a paleopathology of the skull. Ph.D. dissertation, Cambridge University.
1987. Porotic hyperostosis: New evidence to support the anemia theory. *American Journal of Physical Anthropology* 74: 521–6.
- Stuart-Macadam, Patty. 1992. Porotic hyperostosis: A new perspective. *Journal of Physical Anthropology* 87: 39–47.
- Subar, Amy, and Jean Bowering. 1988. The contribution of enrichment and fortification to nutrient intake of women. *Journal of the American Dietetic Association* 88: 1237–45.
- Sullivan, Jerome. 1989. The iron paradigm of ischemic heart disease. *American Heart Journal* 117: 1177–88.
- Taylor, James, and Abraham Bergman. 1989. Iron-fortified formulas: Pediatricians' prescribing practices. *Clinical Pediatrics* 28: 73–5.
- Taylor, John. 1977. Diet, health and economy: Some consequences of planned social change in an aboriginal community. In *Aborigines and change: Australia in the '70s*, ed. R. M. Berndt, 147–58. Canberra.
- Thiele, M., M. Geddes, E. Nobmann, and K. Petersen. 1988. High prevalence of iron deficiency anemia among Alaskan native children. *Journal of the American Medical Association* 259: 2532.
- Thomas, Anita, Valda Bunker, Maureen Stansfield, et al. 1989. Iron status of hospitalized and housebound elderly people: Dietary intake, metabolic balances, hematological and biochemical indices. *Quarterly Journal of Medicine* 70: 175–84.
- Tunnessen, Walter, and Frank Oski. 1987. Consequences of starting whole cow milk at 6 months of age. *Journal of Pediatrics* 111: 813–16.
- Ubelaker, Douglas. 1992. Porotic hyperostosis in prehistoric Ecuador. In *Diet, demography, and disease: Changing perspectives on anemia*, ed. Patricia Stuart-Macadam and Susan Kent, 201–18. New York.
- Volkman, Toby Alice. 1982. The San in transition, Vol I. *Cultural Survival* 9: 1–56.
- Vreugdenhil, G., A. J. Swaak, G. J. Kontoghorghes, and H. G. van Eijk. 1989. Efficacy and safety of oral iron chelator L in anaemic rheumatoid arthritis patients. *Lancet* 2: 1398–9.
- Vyas, Devhuti, and R. K. Chandra. 1984. Functional implications of iron deficiency. In *Iron nutrition in infancy and childhood*, ed. A. Stekel, 45–59. New York.
- Wadsworth, George R. 1992. Physiological, pathological and dietary influences on the hemoglobin level in circulating blood and the influence of iron in the diet. In *Diet, demography, and disease: Changing views of anemia*, ed. Patricia Stuart-Macadam and Susan Kent, 63–104. New York.
- Walker, Phillip. 1986. Porotic hyperostosis in a marine-dependent California Indian population. *American Journal of Physical Anthropology* 69: 345–54.
- Walter, Tomas, Sonia Arredondo, Marianela Arevalo, and Abraham Stekel. 1986. Effect of iron therapy on phagocytosis and bactericidal activity in neutrophils of iron-deficient infants. *American Journal of Clinical Nutrition* 44: 877–82.
- Walter, Tomas, Isidora de Andraca, Patricia Chadud, and Carmen Perales. 1989. Iron deficiency anemia: Adverse effects on infant psychomotor development. *Pediatrics* 84: 7–17.
- Wardrop, C. A. 1987. Runners' anaemia: A paper tiger. *British Medical Journal* 295: 455–6.
- Webb, Stephan. 1989. *Prehistoric stress in Australian aborigines: A paleopathological study of hunter-gatherer population*. BAR International Series 490: 1–197.
- Weinberg, Eugene. 1974. Iron and susceptibility to infectious disease. *Science* 184: 952–6.
1978. Iron and infection. *Microbiological Review* 42: 45–66.
1983. Iron in neoplastic disease. *Nutrition and Cancer* 4: 223–3.
1984. Iron withholding: A defense against infection and neoplasia. *Physiological Reviews* 64: 65–102.
1986. Microorganism. *McGraw-Hill Yearbook of Science and Technology*, 282–3.
1987. Pregnancy-associated immune suppression: Risks and mechanisms. *Microbial Pathogenesis* 3: 393–7.
1989. Cellular regulation of iron assimilation. *Quarterly Review of Biology* 64: 261–90.
1990. Cellular iron metabolism in health and disease. *Drug Metabolism Reviews* 22: 531–79.
1992. Iron withholding in prevention of disease. In *Diet, demography, and disease: Changing perspectives of anemia*, ed. Patricia Stuart-Macadam and Susan Kent, 105–150.
- Witte, David L., Debbie S. Angstadt, Sue H. Davis, and Robert D. Schrantz. 1988. Predicting bone marrow iron stores in anemic patients in a community hospital using ferritin and erythrocyte sedimentation rate. *American Journal of Clinical Pathology* 90: 85–7.
- Woods, S., T. DeMarco, and M. Friedland. 1990. Iron metabolism. *American Journal of Gastroenterology* 85: 1–8.
- Worthington-Roberts, Bonnie, Maryann Breskin, and Elaine Monsen. 1988. Iron status of premenopausal women in a university community and its relationship to habitual dietary sources of protein. *American Journal of Clinical Nutrition* 47: 275–9.
- Yip, Ray, Nancy Binkin, Lee Fleshood, Frederick Trowbridge. 1987. Declining prevalence of anemia among low-income children in the United States. *Journal of American Medical Association* 258: 1619–23.
- Yip, Ray, and Peter Dallman. 1988. The roles of inflammation and iron deficiency as causes of anemia. *American Journal of Clinical Nutrition* 48: 1295–1300.
- Yip, Ray, Samuel Schwartz, and Amos Deinard. 1984. Hematocrit values in white, black, and American Indian children with comparable iron status. *American Journal of Diseases in Childhood* 138: 824–7.
- Youdim, Moussa. 1989. Comment on Pollitt et al. *American Journal of Clinical Nutrition* 50: 687–97.
- Zanella, Alberto, Linda Gridelli, Alessandra Berzuini, et al. 1989. Sensitivity and predictive value of serum ferritin and free erythrocyte protoporphyrin for iron deficiency. *Journal of Laboratory and Clinical Medicine* 113: 73–8.
- Zauber, N. Peter and Ann Zauber. 1987. Hematologic data of healthy very old people. *Journal of American Medical Association* 257: 2181–4.

IV.D.3 Keshan Disease

Discovery and Characteristics of the Disease

Keshan disease (KD) is a unique endemic cardiomyopathy in China with high incidence and mortality. Its etiology and pathogenesis are not as yet completely clear.

In the winter of 1935, an outbreak of an unknown disease with sudden onset of precordial oppression, pain, nausea, vomiting (yellowish fluid), and fatal termination in severe cases occurred in Keshan County, in Heilongjiang Province of northern China. Because its cause was not known, it was named after the place of outbreak by a Japanese military surgeon (Apei 1937).

Later, Keshan disease was also reported from other parts of China and, in fact, research now indicates that the condition has been prevalent in that country for close to 200 years. The earliest-known account of the disease was found in an inscription on a stone pillar at Jinling Temple, Xiaosi village, Huanglong County, Shaanxi Province, in 1812 (Shan and Xue 1987).

Epidemiological Characteristics

There are three major epidemiological characteristics of Keshan disease. The first is its regional distribution. Keshan disease areas are focally distributed in a belt extending from northeast to southwest China and usually located in hilly land. There are isolated spots known as “safety islands” surrounded by affected areas. The second is population susceptibility. Children below 15 years of age and women of childbearing age in northern China, and children below 10 years of age in southern China, constitute the most susceptible populations. They all live in rural areas and in farm families. The third characteristic is seasonal prevalence. The peak season of Keshan disease in northern China is in winter, but in the south it is in summer. There is also a natural fluctuation of prevalence from year to year.

The annual incidence, mortality, and case fatality of Keshan disease in China are shown in Figure IV.D.3.1. The peak years of 1959, 1964, and 1970 had inci-

dences of 6.02, 4.17, and 4.04 per 10,000, respectively, and the mortality was 1.996, 0.978, and 0.676 per 10,000 (Sun et al. 1982; Sun 1987; The Ministry of Public Health of China 1987–95).

Clinical Manifestations

There are no specific symptoms and signs that clearly identify Keshan disease. But according to the degree of heart function insufficiency and of compensative status, Keshan disease is classified into four clinical types: acute, subacute, chronic, and latent (Ge et al. 1983).

Acute cases usually occur suddenly, with acute heart function insufficiency such as cardiogenic shock, severe arrhythmia, and pulmonary edema. The prominent characteristic of the chronic type is chronic congestive heart failure, which is the consequence of acute or subacute types or a result of a long-standing cardiac disorder of an insidious onset. Moderate or severe heart enlargement is always seen in chronic cases.

Subacute cases occur primarily in children. The onset is less sudden and the insidious period is about one or two weeks. The clinical manifestations of the subacute type are mainly cardiogenic shock and/or chronic congestive heart failure, and most patients have facial edema and galloping rhythm of the heart. Latent cases may be discovered as only an incidental finding upon routine physical examination. They usually show a mildly enlarged heart with normal heart function, abnormal electrocardiogram changes of the right bundle branch block, and infrequent premature ventricular contractions.

Pathological Observations

A large number of autopsies have been carried out since the recognition of Keshan disease in 1935. More than 3,600 autopsy cases were collected in the 1950s,

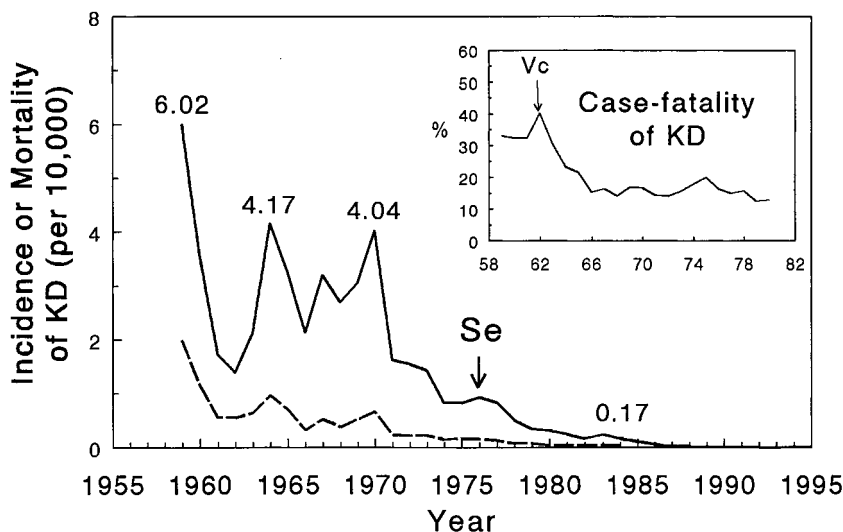


Figure IV.D.3.1. The incidence, mortality, and case-fatality of Keshan disease in China. Solid line (—) is the incidence of Keshan disease (1/10,000). Dashed line (---) is the mortality of Keshan disease (1/10,000).

when there was a heavy prevalence of the disease in northern China. Later, an endemic cardiomyopathy that was prevalent in children in southern China was studied and in 1965 identified as Keshan disease. Since then, large amounts of epidemiological, clinical, and pathological data have been accumulated in the south.

Multifocal necrosis and fibrous replacement of the myocardium are the principal pathological features of Keshan disease (Ge et al. 1983). Two patterns of myocardial necrosis are distinguishable by light microscopy. One is myocytolysis; the other is contraction band necrosis. Myocytolysis exists in the majority of cases and is regarded as a representative lesion of Keshan disease. It seems to be initiated by mitochondrial disorganization and results in the final disappearance of the myofibers. Contraction band necrosis is characterized by myofibril segmentation and has been considered to be the consequence of the severe circulatory disorders predominantly observed in acute cases.

Treatment

There is no specific therapy for Keshan disease. However, the method of administering a megadose of ascorbic acid (vitamin C) in cases of acute and subacute Keshan disease was discovered by Xian Medical College in 1961 (Research Laboratory of Keshan Disease of Xian Medical College 1961). It was a breakthrough in the treatment of Keshan disease and particularly effective in patients with cardiogenic shock. Since then, the case fatality of the acute type of Keshan disease has decreased significantly from 80 percent to less than 20 percent (Figure IV.D.3.1). Patients with congestive heart failure require prompt and optimal digitalization, although the response in some cases is poor. Other treatments (antibiotics, oral diuretics, and moderate restriction of salt) are used for the control of congestive heart failure (Yang et al. 1984).

The Relation between Selenium and Keshan Disease

In the 1960s, it was found that white muscle disease (WMD) in young ruminants caused by selenium (Se) deficiency often occurred in areas affected by Keshan disease, and some similarities in clinical symptoms and pathological features between Keshan disease and white muscle disease were found. The scientists in the Institute of Shaanxi Veterinary Medicine suggested that both Keshan disease and white muscle disease could be responsive to selenium. Small-scale human intervention studies were carried out by several research groups, and some encouraging results were obtained; however, the effectiveness of Se supplementation was demonstrated only by a large-scale selenium intervention study in Mianning County, Sichuan Province, in southwestern China. This study

was carried out from 1974 to 1976 by the Keshan Disease Research Group, Chinese Academy of Medical Sciences (Keshan Disease Research Group, Chinese Academy of Medical Sciences 1979).

Sodium Selenite Intervention

In 1974, 119 production teams observed children of susceptible age (1 to 9 years) in three villages of Mianning County. In 1975, the study was extended to include 169 teams in four villages. One-half of the children were given sodium selenite tablets and the other half a placebo. The assignment to one of these groups was made randomly and remained unchanged during the two years of the investigation. The subjects took sodium selenite once a week, with a dosage of 0.5 milligrams (mg) for those aged 1 to 5 years and a dose of 1.0 mg for those 6 to 9 years old. Because of the convincing results obtained in 1974 and 1975 (to be discussed shortly), the control groups were abandoned and all subjects were given sodium selenite in 1976 and 1977.

A Keshan disease hospital was established in the area under investigation; there, diagnosis and subtyping of the disease were carried out according to the criteria set up in 1974 by the National Seminar of the Etiology of Keshan Disease. Electrocardiograms, heart roentgenograms, and physical examinations were performed on patients admitted to the hospital for observation. In some cases, blood Se content, glutathione peroxidase (GPX), serum glutamic oxalacetic transaminase (GOT), and glutamic pyruvic transaminase (GPT) activity were determined. Patients receiving treatment at other medical units were transferred to the hospital after the improvement of their general condition. Follow-up examinations were carried out each year to study the progress of individual patients.

The incidence and prognosis of the subjects investigated are shown in Table IV.D.3.1. In 1974, among the 3,985 children in the control group, there were 54 cases of Keshan disease (13.55 per 1,000), whereas only 10 of the 4,510 Se-supplemented subjects fell ill (2.22 per 1,000). The difference between the incidence of the two groups was highly significant ($P < 0.01$). A similar difference was found in 1975. In 1976, when all subjects were given selenite, only 4 cases out of 12,579 subjects occurred, which further lowered the incidence to 0.32 per 1,000. There was 1 case of the typical subacute type among 212 children who failed to take selenite. In 1977, there were no new cases among the 12,747 supplemented subjects. These results indicated that sodium selenite intervention had a significant effect in reducing incidence of Keshan disease. The Se-supplemented subjects had not only a lower incidence but also lower mortality and better prognosis.

In 1976 and 1977, liver function tests and general physical examinations were given to 100 subjects who had taken Se tablets weekly for three to four years. The results were not significantly different from

Table IV.D.3.1. *Keshan disease incidence and prognosis of selenium-supplemented and control children (1–9 years old) in Mianning County, Sichuan Province, China, during 1974–7^a*

Groups	Year	Subjects	New cases	Incidence (‰)	Prognosis			
					Turned chronic	Turned latent	Improved	Death
Se-supplemented	1974	4,510	10	2.22	1	9	0	0
	1975	6,767	7	1.03	0	6	0	1
	1976	12,579	4	0.32	0	2	0	2
	1977	12,747	0	0	0	0	0	0
(Total)		36,603	21	0.57	1	17	0	3
Control	1974	3,985	54	13.55	2	16	9	27
	1975	5,445	52	9.55	3	13	10	26
	1976	212	1	4.72	1	0	0	0
(Total)		9,642	107	11.10	6	29	19	53

^aThere are significant differences of incidence between the two groups in each year and in total ($P < 0.01$).

those for the unsupplemented children and indicated that selenium-supplementation produced no untoward side effects.

In 1976, observations on the effects of sodium selenite were extended beyond Mianning County to include Dechang, Xichang, Yuexi, and Puge Counties. All children (1 to 12 years old) in some of the most severely affected areas were supplemented with selenite as just described, whereas the children in the nearby areas served as unsupplemented controls. The results, summarized in Table IV.D.3.2, show that in each year the incidence of Keshan disease among the Se-supplemented children of the five counties was significantly lower than that among the unsupplemented children. Similar results were obtained by Xian Medical College in a study in Huanglong County, Shaanxi Province, from 1975 to 1977 (Research Laboratory of Keshan Disease of Xian Medical College 1979).

Because of the convincing evidence of the efficacy of sodium selenite in preventing Keshan disease, it has been widely used since 1976 in tablet form, as an addition to table salt, and as a fertilizer. Concomitantly, the incidence and mortality of Keshan disease decreased to 0.17 and 0.04 per 10,000, respectively (Figure IV.D.3.1).

Occurrence in Selenium-Deficient Areas

Since regional distribution is the first epidemiological characteristic of Keshan disease, the discovery of environmental differences between areas that are affected by Keshan disease and those that are not has been an important goal in the study of its etiology.

A fluorometric method for determining Se content and another method for determining glutathione peroxidase (GPX) activity were adapted and set up by the Keshan Disease Research Group of the Chinese Academy of Medical Sciences in 1972 and 1975. Sampling techniques indicated that blood and hair Se content and blood GPX activity of residents in Keshan disease-affected areas were lower than those of people in nonaffected areas (see Table IV.D.3.3). However, after the oral administration of sodium selenite for 1 year, the blood GPX activity of children in Keshan disease-affected areas had increased to levels comparable to those of children in nonaffected areas (see Table IV.D.3.4).

It was assumed that the low-selenium status of people living in Keshan disease-affected areas was caused by a low dietary selenium intake. Therefore, large samples of cereals, soil, water, and human blood, hair, and urine from Keshan disease-affected areas (11 provinces, 42 counties, 77 spots) and nonaffected

Table IV.D.3.2. *Keshan disease incidence in selenium-supplemented and control children (1–12 years old) in five counties of Sichuan Province, China, during 1976–80^a*

Year	Se-supplemented			Control		
	Subjects	New cases	Incidence (‰)	Subjects	New cases	Incidence (‰)
1976	45,515	8	0.17	243,649	488	2.00
1977	67,754	15	0.22	222,944	350	1.57
1978	65,953	10	0.15	220,599	373	1.69
1979	69,910	33	0.47	223,280	300	1.34
1980	74,740	22	0.29	197,096	202	1.02
Total	323,872	88	0.27	1,107,568	1,713	1.55

^aThere are significant differences of incidence between the two groups in each year and in total ($P < 0.01$).

Table IV.D.3.3. *Selenium levels in human blood and hair from residents in Keshan disease-affected and nonaffected areas in 1972-3*

Subjects ^a	Blood Se (ng/g) ^b	Hair Se (ng/g) ^b
KD-affected areas		
KD patients	18 ± 9 (42) ^a	107 ± 53 (21) ^f
Farmers	21 ± 7 (24) ^a	
Nonaffected areas (near KD areas)		
Farmers	32 ± 4 (20) ^b	187 ± 51 (20) ^g
Staff members in town	41 ± 21 (10) ^c	
Nonaffected areas (far from KD areas)		
Staff members (A) in city	173 ± 44 (16) ^d	712 ± 97 (13) ^h
Staff members (B) in city	255 ± 38 (17) ^e	834 ± 156 (11) ^k

^aStaff members (A) were the members of a KD research group of the Chinese Academy of Medical Sciences and had been in KD areas for less than one year; staff members (B) had never been in KD areas.

^bValues are means ± SD; number of samples is in parentheses; values not sharing a common superscript in a vertical column are significantly different ($P < 0.05$) by the t-test.

areas (20 provinces, 86 counties, 110 spots) were collected. The data indicated that the samples containing less than 20 nanograms/gram (ng/g) of Se in blood or 120 ng/g in hair or 10 ng/g in cereals were almost exclusively from Keshan disease-affected areas, whereas samples with Se content of more than 50 ng/g in blood or 200 ng/g in hair or 20 ng/g in cereals were all from nonaffected areas (Yang et al. 1982). These data could be considered the threshold of Se content in assessing the risk of Keshan disease, especially for those who live on a rather simple diet composed mainly of locally grown cereals (Table IV.D.3.5).

Similar results from more than 10,000 samples were reported by the Institute of Geography, Chinese Academy of Sciences (Group of Environmental and Endemic Disease 1982; Xu et al. 1982). Based on a large amount of data, a map of selenium distribution in China was published in *The Atlas of Endemic Diseases and Their Environments in the People's*

Table IV.D.3.4. *Blood glutathione peroxidase (GPX) activities of children from Keshan disease-affected and nonaffected areas in 1975*

Subjects	Blood GPX (units)
KD-affected areas	
KD child patients	57.1 ± 6.1 (22) ^a
Healthy children	60.5 ± 5.6 (63) ^a
Se-supplemented children	76.1 ± 8.4 (58) ^b
Nonaffected areas (near KD areas)	
Children in farmer's family	73.6 ± 13.8 (20) ^b
Nonaffected areas (far from KD areas)	
Children in farmer's family	77.5 ± 9.8 (22) ^b

Note: Values are means ± SD; number of samples is in parentheses; values not sharing a common superscript are significantly different ($P < 0.01$) by the t-test.

Table IV.D.3.5. *Selenium contents of blood, hair, and grains in Keshan disease-affected and nonaffected areas*

Samples	KD areas (ng/g)	Overlapping areas (ng/g)	Non-KD areas (ng/g)
Human blood	< 20	20-50	> 50
Human hair	<120	120-200	>200
Staple grain	< 10	10-20	> 20

Republic of China (1989). It shows that low-selenium areas are distributed in a belt extending from northeast to southwest China, which is consistent with the geographical distribution of Keshan disease.

The results obtained by medical, veterinary, and geographical scientists all indicated that Keshan disease occurs only in selenium-deficient areas, which explains the epidemiological characteristic of its regional distribution. The fact that the blood and hair Se levels and GPX activity of farming residents are lower than those of nonfarming residents in the same Keshan disease-affected areas at least partially explains the characteristic of population susceptibility. In view of the results just described, it was concluded that selenium deficiency is the basic cause of Keshan disease occurrence.

Biochemical Changes

There were no significant differences in Se status and GPX activity between Keshan disease patients and other residents of the same areas (Zhu, Xia, and Yang 1982). There were also no changes in hair Se content to accompany the seasonal or yearly fluctuation of Keshan disease prevalence (Sun et al. 1980). These facts suggest that the etiology of Keshan disease cannot be explained by selenium deficiency alone and that there is a need to understand the ways in which selenium prevents Keshan disease.

A Study on Keshan Disease Patients

A comprehensive scientific survey of Keshan disease in Chuxiong prefecture, Yunnan Province, southwestern China, was conducted from 1984 to 1986 by 293 workers from 16 laboratories in seven provinces. The objective was to examine 3,648 children from 56 villages in Keshan disease-affected and nonaffected areas and compare the two groups. One hundred and sixty-seven children who were Keshan disease patients were treated, and 27 autopsies were carried out on patients who had died of subacute Keshan disease. Autopsy controls consisted of those who died from other diseases in both affected and nonaffected areas. The results of this survey were published in the *Collected Work of Comprehensive Scientific Survey on Keshan Disease in Chuxiong Prefecture (1984-1986)* (Yu 1988).

Table IV.D.3.6 *Selenium contents and glutathione peroxidase (GPX) activities in tissues from patients with subacute Keshan disease (Sub-KD) and controls in affected (Control-A) or nonaffected areas (Control-NA)*

		Sub-KD	Control-A	Control-NA
Selenium (ng/g fresh)	Heart	25 ± 10 (6) ^a	55 ± 32 (4) ^a	134 ± 32 (3) ^b
	Liver	52 ± 30 (6) ^a	89 ± 50 (4) ^a	256 (1)
GPX (U/mg protein)	Heart (H ₂ O ₂)	10 ± 5 (11) ^a	26 ± 21 (7) ^b	73 ± 12 (20) ^c
	(t-BOOH)	10 ± 5 (11) ^a	29 ± 25 (7) ^b	80 ± 12 (2) ^c
	Liver (H ₂ O ₂)	7 ± 4 (4) ^a	38 ± 11 (3) ^b	
	(t-BOOH)	23 ± 5 (4) ^a	47 ± 25 (3) ^b	
	Kidney (H ₂ O ₂)	20 ± 7 (4) ^a	78 ± 39 (3) ^b	
	(t-BOOH)	32 ± 2 (4) ^a	85 ± 36 (3) ^b	

Note: Values are means ± SD; number of samples is in parentheses; values not sharing a common superscript in a horizontal row are significantly different ($P < 0.01$) by the t-test.

The comparison of Se content and GPX activity in the tissues of patients with subacute Keshan disease (Sub-KD) and those of the controls in affected (Control-A) and nonaffected areas (Control-NA) is shown in Table IV.D.3.6. There was a tendency for Keshan disease patients to have lower heart and liver Se levels than the controls, and their GPX activity was significantly lower than that of the controls. Other results supported the hypothesis that the heart is the organ most susceptible to selenium deficiency.

Other oxidant defense indices in the myocardium of child patients with Sub-KD were measured. The results showed that the activity of superoxide dismutase (SOD) in the myocardia of Sub-KD patients was significantly lower than that of Control-NA subjects, but malondialdehyde (MDA) and free radicals were higher than those of Control-NA. These results indicated that there was an accumulation of lipid peroxides and free radicals in Keshan disease patients. However, there were no differences between Sub-KD patients and controls in glutathione reductase (GR) and glutathione-S-transferase (GST) activity.

Ultrastructural observations showed that mitochondria in the myocardia of Sub-KD patients appeared to be most commonly and conspicuously affected and were involved early. In general, the changes in biochemical functions occurred before the structural damage. The results indicated that the activity of succinate dehydrogenase (SDH), cytochrome C oxidase (CCO), and H⁺-ATPase, and content of Co Q in the myocardial mitochondria of Sub-KD patients were lower than those of Control-NA. These findings implied that the lesions of the mitochondria affect mostly the respiratory chain of the inner membrane. Changes in membrane potential and decreased fluidity of membrane lipid were observed in the myocardial mitochondria of Sub-KD patients. The content of Se in the mitochondria was one-eighth that of Control-NA (Yu 1988). It is likely that the structural and functional abnormalities of the myocardial mitochondria of patients with Keshan disease may result from selenium deficiency. In addition, the mitochondria constitute one of the calcium pools and play an important role in regulating the calcium concentration in cells.

The increased calcium content in the myocardial mitochondria of Sub-KD patients may have an intrinsic relationship with the functional abnormalities.

From this comprehensive survey, it was concluded that relatively weak oxidant defenses, mainly low GPX activity, and insufficient vitamin E (less than 8 µg/ml plasma in people living in Keshan disease-affected areas) created a selenium-deficient population susceptible to oxidant stresses that resulted in damage to myocardial structures and functions and finally in the occurrence of Keshan disease.

Effects of Se Deficiency

It was not clear whether differences in dietary patterns (aside from selenium intake) between people in Keshan disease-affected and nonaffected areas were involved in the etiology of Keshan disease. Three studies were carried out in 1987, 1989, and 1991 in Mianning and Dechang Counties, Sichuan Province, where there was a very high incidence of Keshan disease there was during the 1970s. These two counties are only 150 kilometers (km) apart and have the same dietary patterns. According to the results from a nutrition survey in 1985, conducted by the Anti-Epidemic Station of Liangshan Autonomous Prefecture, the dietary selenium intake was 11 µg per day for both counties. The only difference was a supplementation of selenized salt (15 µg selenium per g salt, as sodium selenite) in Mianning County from 1983 onward, so the people of Mianning had 69 µg of extra selenium per day, giving them a total selenium intake of 80 µg per day. Table IV.D.3.7 shows the results of a study conducted in 1987 (Xia, Hill, and Burk 1989; Hill et al. 1996) following selenized salt supplementation for 4 years. Not only were the blood Se levels of children in Mianning County higher than those in Dechang County, but the GPX activity of their plasma was three times as high, and plasma selenoprotein P concentrations were also much higher. Linear regression analysis of plasma showed that selenium concentration, GPX activity, and selenoprotein P concentration correlated well with each other. There were no differences in other indices for oxidant defense capability between the residents of the two counties.

Table IV.D.3.7. *Indexes for oxidant defense capability in the blood of children from Dechang and Mianning Counties in 1987*

		Dechang (-Se)		Mianning (Se-salt)	
Selenium	Blood (ng/g)	18 ± 10	(22) ^a	47 ± 9	(18) ^b
	Plasma (ng/g)	13 ± 55	(22) ^a	40 ± 11	(18) ^b
	RBC (ng/g)	26 ± 9	(22) ^a	47 ± 9	(19) ^b
GPX	Plasma (U/L)	29 ± 15	(21) ^a	87 ± 15	(15) ^b
	RBC (U/gHb)	2.8 ± 1.7	(21) ^a	8.4 ± 3.3	(19) ^b
Selenoprotein P	Plasma (U/L)	0.10 ± 0.04	(22) ^a	0.39 ± 0.17	(17) ^b
SOD	RBC (U/mgHb)	17 ± 7	(22)	17 ± 3	(19)
Catalase	RBC (U/gHb)	238 ± 48	(22)	231 ± 29	(19)
Glutathione reductase	RBC (U/gHb)	3.3 ± 1.0	(22)	4.2 ± 1.3	(19)
GSH	Plasma (μM)	4.7 ± 0.6	(21)	4.4 ± 1.0	(16)
	RBC (mM)	2.5 ± 0.4	(19)	2.1 ± 0.7	(19)
Vitamin E	Plasma (μg/ml)	4.0 ± 1.2	(22)	3.8 ± 1.2	(19)
MDA	Plasma (nmol/ml)	6.7 ± 2.2	(22)	6.0 ± 2.6	(19)

Note: Values are means ± SD; number of samples is in parentheses; values not sharing a common superscript in a horizontal row are significantly different ($P < 0.001$) by the t-test.

Effects of age and sex. In the 1989 study, blood samples were taken from 401 healthy subjects (half male and half female) in Dechang County. The lowest levels of selenium and GPX activity were found in the youngest (2 to 5 years old) and oldest (more than 60 years old), and no significant sex difference was found. These data were consistent with the occurrence of Keshan disease in children, the susceptible population group (Xia et al. 1994).

Changes in selenium intake. It was found that the selenium status and GPX activity of residents of Dechang County increased gradually year after year to reach the levels of persons in nonaffected areas in 1991. Coincidentally, there was no occurrence of Keshan disease in this year. Therefore, it was necessary to check the selenium intake of the local residents. In the 1991 study, the selenium contents of rice, table salt, and the mixed diet were measured. The data showed that selenium in rice was $16.6 \pm 4.3 \mu\text{g/kg}$ ($n = 10$), which was in the same range as that of KD-affected areas ($<20 \text{ ng/g}$ in grain, see Table IV.D.3.5), but selenium intake from the diet was estimated to be $38 \pm 28 \mu\text{g/day}$ ($n = 30$), which reached the minimum requirement of $20 \mu\text{g/day}$ (Yang et al. 1985; Yang and Xia 1995). This change probably resulted from the inadvertent distribution of selenized salt in Dechang County and, with the improvement of the local economy, the increased consumption of food from selenium-adequate areas. It was confirmed that the decrease of Keshan disease incidence in Dechang County correlated with an improved selenium status because of an increase in dietary selenium intake.

Marginal vitamin E status. Vitamin E (VE), an important constituent of the oxidant defense system, protects biological membrane tissue as a free-radical scavenger and has a synergistic effect with selenium.

Animal studies have indicated that cardiac injury can be developed in animals fed a diet deficient in both selenium and vitamin E but that neither deficiency by itself causes cardiac injury (Van Vleet, Ferrans, and Ruth 1977; Konz et al. 1991). Other animal studies have indicated that myocardial injury can occur in animals fed a diet of foods from Keshan disease-affected areas and that this condition can be alleviated by the supplementation of selenium or vitamin E or both (Wang et al. 1991). Plasma vitamin E of residents of Keshan disease-affected areas ranged from 2 to $8 \mu\text{g/ml}$, which is a low marginal status (Keshan Disease Research Group, Chinese Academy of Medical Sciences 1977a; Li et al. 1988; Xia et al. 1989; Zhu et al. 1991). This low marginal vitamin-E status could act as a promoting cofactor in the occurrence of Keshan disease. If so, vitamin E supplementation might be helpful in improving the low selenium status of residents of Keshan disease areas.

The Etiology of Keshan Disease

Although the overall incidence of Keshan disease has steadily declined since the 1980s, there was a question of whether there was a corresponding change in the selenium status of local populations at risk. Geographic scientists collected 4,600 samples (rock, soil, water, cereals, animal fur, and human hair) from 217 sampling spots and monitored the incidence of Keshan disease for three years in the 1980s (Tang et al. 1991). The results (see Table IV.D.3.8) indicated that the environmental selenium status was still as low as that in the 1970s, but the selenium level in hair had increased to a normal level (see Table IV.D.3.5). It is believed that in addition to deliberate selenium supplementation, there has been inadvertent supplementation as a previously self-sufficient lifestyle gave way to a commodity economy, in which foods with higher selenium content were imported into the affected

Table IV.D.3.8. Comparison of selenium contents in cereals and human hair between the 1970s and 1980s

	1970s	1980s
Corn (µg/kg)	13 ± 6 (157) ^a	10 ± 9 (90) ^b
Rice (µg/kg)	14 ± 8 (47)	15 ± 11 (31)
Wheat (µg/kg)	14 ± 7 (115)	13 ± 9 (96)
Hair (µg/kg)	85 ± 32 (815) ^a	192 ± 96 (44) ^b

Note: Values are means ± SD; number of samples is in parentheses; values not sharing a common superscript in a horizontal row are significantly different ($P < 0.001$) by the t-test.

areas. In addition there has been an increase in the consumption of animal products in recent years. All of this would seem to confirm that an adequate selenium status is necessary for the prevention of Keshan disease.

Although selenium deficiency can explain the epidemiological characteristic of the regional distribution of Keshan disease, it only partially explains the susceptibility of populations. Because the selenium status of residents in Keshan disease-affected areas does not change with seasonal fluctuation of the incidence of the disease, another potential cofactor (or cofactors) must be involved in its occurrence. Such cofactors would act as promoters or stimulators during selenium deficiency.

The potential cofactors in the etiology of Keshan disease can be classified into four categories:

1. An insufficiency of nutrients that have relative biological function with selenium, such as VE, methionine, and iodine.
2. An excess of elements that have antagonistic interaction with selenium, leading to poor availability of selenium.
3. The prevalence of mycotoxins that contaminate foods and damage the myocardium.
4. A virus infection in which the target organ is the heart (Su 1979; Bai et al. 1984).

Present information does not permit ruling out any one of these as a cofactor in selenium deficiency and in the etiology of Keshan disease.

Clearly, much work remains to be done, even though studying the etiology of Keshan disease presents some difficulties. One problem is that there are no ideal animal models for Keshan disease; another is that almost no new Keshan disease patients can be found (only 57 new patients were reported in the whole country in 1995). However, it is believed that the pathogenic factors in Keshan disease areas are still active because, according to the Annual Report from the Ministry of Public Health of China there are about 50,000 latent Keshan disease patients in the country. Therefore, work will continue.

Yiming Xia

Bibliography

- Apei, H. 1937. Report of etiological survey on Keshan disease. *Research Reports of Continental Academy* 1: 1-10.
- Arthur, J. R., F. Nicol, and G. J. Beckett. 1990. Hepatic iodothyronine deiodinase: The role of selenium. *Biochemical Journal* 272: 537-40.
- Bai, J., S. Wu, K. Ge, et al. 1984. The combined effect of selenium deficiency and viral infection on the myocardium of mice. *Acta Academiae Medicinae Sinica* 2: 29-31.
- Beck, M. A., P. C. Kolbeck, L. H. Rohr, et al. 1994a. Amyocarditic coxsackievirus becomes myocarditic in selenium deficient mice. *Journal of Medical Virology* 43: 166-70.
- 1994b. Vitamin E deficiency intensifies the myocardial injury of coxsackievirus B3 infection of mice. *Journal of Nutrition* 124: 345-58.
- Beck, M. A., P. C. Kolbeck, Q. Shi, et al. 1994. Increased virulence of a human enterovirus (coxsackievirus B3) in selenium deficient mice. *The Journal of Infectious Diseases* 170: 351-7.
- Beck, M. A., Q. Shi, V. C. Virginia, and O. A. Levander. 1995. Rapid genomic evolution of non-virulent coxsackievirus B3 in selenium-deficient mice results in selection of identical virulent isolates. *Nature Medicine* 1: 433-6.
- Becker, K. B., M. J. Schneider, J. C. Davey, and V. A. Galton. 1995. The type III 5-deiodinase in *Rana catesbeiana* tadpoles is encoded by a thyroid hormone-responsive gene. *Endocrinology* 136: 4424-6.
- Behne, D., A. Kyriakopoulos, H. Meinhold, and J. Köhrle. 1990. Identification of type I iodothyronine 5'-deiodinase as a selenoenzyme. *Biochemical and Biophysical Research Communications* 173: 1143-9.
- Davey, J. C., K. B. Becker, M. J. Schneider, et al. 1995. Cloning of a cDNA for the type II iodothyronine deiodinase. *Journal Biological Chemistry* 270: 26786-9.
- Editorial Board of *The Atlas of Endemic Diseases and Their Environments in the People's Republic of China*. 1989. *The atlas of endemic diseases and their environments in the People's Republic of China*. Beijing.
- Ge, K., A. Xue, J. Bai, and S. Wang. 1983. Keshan disease - an endemic cardiomyopathy in China. *Virchows Archive of Pathological Anatomy and Histopathology* 401: 1-15.
- Ge, K., A. Xue, and S. Wang. 1980. Ultrastructural changes of myocardial necrosis of Keshan disease. Light and electron microscope comparative study with a cytolocalization technique. *Chinese Journal of Medicine* 60: 407-9.
- Group of Environmental and Endemic Disease. 1982. The relation of Keshan disease to the natural environment and the background of selenium nutrition. *Acta Nutrimenta Sinica* 4: 175-82.
- Hill, K. E., and R. F. Burk. 1982. Effect of selenium deficiency and vitamin E deficiency on glutathione metabolism in isolated rat hepatocytes. *Journal of Biological Chemistry* 257: 10668-72.
- Hill, K. E., and R. F. Burk. 1994. Selenoprotein P - an extracellular protein containing multiple selenocysteines. In *Selenium in biology and human health*, ed. R. F. Burk, 119-31. New York.
- Hill, K. E., Y. Xia, B. Kesson, et al. 1996. Selenoprotein P concentration in plasma is an index of selenium status in selenium-deficient and selenium-supplemented Chinese subjects. *Journal of Nutrition* 126: 138-45.
- Keshan Disease Research Group, Chinese Academy of Medical Sciences. 1974. A study on the relationship between selenium and Keshan occurrence. In *The Proceedings*

- of the National Symposium on the Etiology of Keshan Disease, 181-200. Shenyang.
- 1977a. Comparison of vitamin E contents of serum from residents living in Keshan disease affected and non-affected areas. *Collected papers on Keshan disease* 1: 71-7. Beijing.
- 1977b. Glutathione peroxidase activities in blood of residents in Keshan disease areas. *Collected papers on Keshan disease* 1: 67-70.
1979. Observation on the effect of sodium selenite in prevention of Keshan disease. *Chinese Medical Journal* 92: 471-6.
- Konz, K. H., R. A. Walsh, K. E. Hill, et al. 1991. Cardiac injury caused by selenium and vitamin E deficiency in the minipig: Biochemical and hemodynamic studies. *Journal of Trace Elements in Experimental Medicine* 4: 61-7.
- Li, G., J. Yang, Y. Ren, et al. 1988. Distribution of lipid metabolism in children in endemic areas of Keshan disease. In *Collected work of comprehensive scientific survey on Keshan Disease in Chuxiong Prefecture 1984-1986*, ed. W. Yu, 247-9. Beijing.
- Ministry of Public Health of China. 1987-95. Annual reports on national healthy statistics. Beijing.
- Research Laboratory of Keshan Disease, Xian Medical College. 1961. Study on clinical effect and mechanism in treatment of acute Keshan disease with large dosage of vitamin C. *Chinese Journal of Internal Medicine* 6: 346-50.
1979. Observations on the efficacy of sodium selenite in the prevention of acute onset of Keshan disease. *Chinese Journal of Internal Medicine* 59: 457-9.
- Shan, S., and A. Xue. 1987. The information of Keshan disease and its prevention and treatment in Shaanxi Province. In *The study on prevention and treatment of Keshan disease in China*, ed. W. Yu, 61-8. Beijing.
- Su, C. 1979. Preliminary results of viral etiology of Keshan disease. *Chinese Medical Journal* 59: 466-72.
- Sun, J. 1987. The policy of "the prevention first" is most important in the prevention and treatment of Keshan disease - review on the history of preventing and treating Keshan disease in China. In *The study on prevention and treatment of Keshan disease in China*, ed. W. Yu, 1-24. Beijing.
- Sun, J., Y. Lu, G. Wu, and R. Teng. 1982. A survey of the epidemic of Keshan disease in the north of China. *Chinese Journal of Endemiology* 1: 2-5.
- Sun, S., T. Yin, G. Wang, et al. 1980. The relationship between seasonal prevalence of Keshan disease and hair selenium of inhabitants. *Chinese Journal of Preventive Medicine* 14: 17-9.
- Tang, J., W. Zhu, R. Li, and W. Wang. 1991. The relationship between Keshan disease and environment selenium, some vital elements. *Chinese Journal of Endemiology* 10: 269-74.
- Ursini, F., M. Maiorino, and C. Gregolin. 1985. The selenoenzyme phospholipid hydroperoxide glutathione peroxidase. *Chinese Journal of Biochemistry and Biophysics* 839: 62-70.
- Van Vleet, J. F., V. J. Ferrans, and G. R. Ruth. 1977. Ultrastructural alterations in nutritional cardiomyopathy of selenium-vitamin E deficient swine. I. Fiber lesions. *Laboratory Investigations* 37: 188-96.
- Wang, F., D. Kang, S. Liu, et al. 1991. The relation of both selenium and vitamin E deficiency to etiology of Keshan disease. *Chinese Journal of Control of Endemic Disease* 6: 306-7.
- Xia, Y., K. E. Hill, and R. F. Burk. 1989. Biochemical studies of a selenium-deficient population in China: Measurement of selenium, glutathione peroxidase and other oxidant defense indices in blood. *Journal of Nutrition* 119: 1318-26.
- Xia, Y., J. Piao, K. E. Hill, and R. F. Burk. 1994. Keshan disease and selenium status of populations in China. In *Selenium in biology and human health*, ed. R. F. Burk, 181-96. New York.
- Xu, G., W. Xue, P. Zhang, et al. 1982. Selenium status and dietary selenium content of populations in the endemic and non-endemic areas of Keshan disease. *Acta Nutrimenta Sinica* 4: 183-90.
- Yang, G., J. Chen, Z. Wen, et al. 1984. The role of selenium in Keshan disease. In *Advances in nutrition research*, Vol. 6, ed. H. H. Draper, 203-31. New York.
- Yang, G., G. Wang, T. Yin, et al. 1982. Relationship between the distribution of Keshan disease and selenium status. *Acta Nutrimenta Sinica* 4: 191-200.
- Yang, G., and Y. Xia. 1995. Studies on human dietary requirements and safe range of dietary intakes of selenium in China and their application in the prevention of related endemic diseases. *Biomedical and Environmental Sciences* 8: 187-201.
- Yang, G., R. Zhou, S. Yin, et al. 1985. Investigation on selenium requirements of Chinese people. I. Physiological, minimal requirements and minimum dietary allowance. *Hygiene Investigation* 14: 24-8.
- Yu, W. 1988. *Collected work of comprehensive scientific survey on Keshan disease in Chuxiong prefecture (1984-1986)*. Beijing.
- Zhu, L., Y. Cheng, J. Piao, et al. 1991. The antioxidative ability and oxidative factors of children in Keshan disease affected areas of Sichuan Province. *Chinese Journal of Endemiology* 10: 7-11.
- Zhu, L., Y. Xia, and G. Yang. 1982. Blood glutathione peroxidase activities of populations in Keshan disease affected and non-affected areas. *Acta Nutrimenta Sinica* 4: 229-33.

IV.D.4 ☞ Osteoporosis

Calcium and Life

Calcium is the fifth most abundant element in the biosphere, after oxygen, silicon, aluminum, and iron. It is present in high concentration in seawater and in all fresh waters that support an abundant biota. Fortunately, the calcium ion has just the right radius to fit neatly within the folds of various peptide chains. Calcium thereby stabilizes and activates a large number of structural and catalytic proteins essential for life. In this capacity calcium serves as a ubiquitous second messenger within cells, mediating such diverse processes as mitosis, muscle contraction, glandular secretion, blood coagulation, and interneuronal signal transmission. Controlling these activities requires careful regulation of the concentration of calcium in critical fluid compartments. This regulation is accomplished in two basic ways.

At a cellular level, calcium is ordinarily sequestered

within intracellular storage compartments. It is released into the cell sap when needed to trigger various cellular activities, and then quickly pumped back into its storage reservoirs when the activity needs to be terminated. This control mode is exemplified by the accumulation and release of calcium by the sarcoplasmic reticulum of striated muscle. The second type of control, utilized by many tissues in higher organisms, is the tight regulation of the calcium level in the blood and extracellular fluids that bathe all the tissues. Individual cells, needing a pulse of calcium, simply open membrane channels and let calcium pour in from the bathing fluid; they then pump it back out when the particular activity needs to cease.

Bone and the Regulation of Calcium Levels

Each mode of control requires both a reserve supply of calcium and a place to put an excess of calcium: In the first mode, the source and sink are within the cell; and in the second, they are outside the cell but still within the organism. The extracellular calcium reserve (and sink) in the higher vertebrates has, over the course of evolution, become the organ system we call bone. Along the way, building on the hardness of calcium deposits, bone acquired the mechanical and structural functions that have become its most prominent features.

The fossil record shows that bone evolved independently many times over the course of vertebrate evolution, usually in a marine environment where bone probably functioned primarily as a sink for calcium (since the fluid in contact with the gill surfaces represented an essentially inexhaustible source). The hardness of bone served many useful, but secondary purposes, ranging from dermal armor, to teeth, to internal stiffening. As mechanisms of controlling concentration of minerals in the internal environment evolved to higher levels of refinement and an internal sink became less necessary, the internal skeleton dropped out of many fish genera, which retained only the structural portions that were vital – teeth and dermal armor.

But in amphibians and terrestrial vertebrates, living outside of a buoyant medium, the internal stiffening could not be dropped. It provided structural support and mechanical strength, and it permitted movement against gravity. Also, deprived of constant contact with a bathing medium high in calcium, the organism now became more dependent upon internal reserves of calcium to ensure maintenance of constant calcium concentrations in the extracellular fluids.

While this need for a calcium reserve in terrestrial vertebrates is virtually self-evident, it is useful to note that the sink function, sequestering of excess calcium, remains important on dry land as well. (If calcium were constantly in short supply, of course, then a reserve sufficient to serve a structural function could never be accumulated in the first place.) Typically, in the life of a terrestrial vertebrate, the reserve function

of the skeleton is called upon only intermittently. At other times the skeleton stores excesses of calcium made available from the environment. This process, as already noted, is constantly needed in a marine habitat but occurs mainly on feeding for most terrestrial vertebrates.

Once vertebrates came out onto dry land, two sometimes competing objectives had to be managed: maintaining the extracellular fluid calcium level and maintaining the size of the skeletal reserve. Whereas the former could be managed in a marine environment by adjusting fluxes of ions across the gill membranes, on dry land it had to be done by adjusting the net flow of calcium into and out of bone. In this process bone mass itself actually changes as the skeleton functions to support the body fluid calcium levels, and skeletal structural strength necessarily changes in parallel. The size of the reserve, which is the basis for bone strength, is ultimately limited by forces acting outside the skeleton, that is, by adjusting inflow from ingested foods and outflow through the kidney, as well as by a mechanical feedback system within the skeleton (see following discussion). It must be noted, however, that the extraskeletal portion of that regulatory system works adequately only when ingested food contains sufficient calcium.

Calcium Abundance in the Diets of Terrestrial Vertebrates

Calcium is so abundant in even the terrestrial environment that most wild foods contain relatively large quantities of it. In fact, much of the calcium content of several leafy plants, notably the halophytes (for example, spinach), represents a plant tissue analogue of the bony sink of marine vertebrates. In other words, the plant creates calcium deposits as a means of keeping calcium levels from rising too high in plant tissue fluids. However, that sequestered calcium remains a part of the plant and hence becomes available to the animal eating it.

By the time they eat sufficient food to meet total energy needs, most mammals have inevitably ingested a great deal of calcium. This calcium load is generally so great that higher vertebrates have evolved mechanisms to prevent being swamped by an excess of calcium. One of these has been the development of a relative absorptive barrier at the intestine, and the second has been the ability to damp out any elevations in extracellular fluid calcium by promptly transferring an excess of absorbed calcium into bone.

But there are some delicate trade-offs involved here, particularly in terrestrial vertebrates, where the structural significance of the skeleton is crucial (even if a secondary function, from an evolutionary standpoint). The absorptive barrier can only be partial, or it would not be possible to accumulate adequate skeletal mass to serve a structural function, nor to repay temporary withdrawals from the reserve. Urinary excretion might take care of an absorbed surplus, and

in fact certainly does so in most vertebrates, but calcium is relatively insoluble, and renal capacity for handling large excesses of calcium is limited by the propensity of the kidney tissue to calcify. Thus, organisms evolved mechanisms for temporarily putting excess calcium into the skeleton during the absorptive phase after feeding and then withdrawing it from the skeleton, as needed, during periods of fasting or starvation.

The need for a reserve during fasting is not because calcium is consumed in the various metabolic processes that it activates (as would be the case, for example, with ascorbic acid or the B vitamins) but because calcium is lost every day through skin and excreta. Further, during childbearing, calcium is transferred from the mother to the progeny, both in utero and through lactation. Hence, there is an obligatory calcium need throughout life, first to accumulate skeletal mass during growth and then to offset daily losses at all ages.

Bone Remodeling and Bone Mass

Bone has no capacity simply to soak up or release calcium ions on need. Rather, these functions are served by forming and destroying actual packets of bony tissue. Collectively these processes of formation and resorption are termed "bone remodeling." Remodeling occurs throughout life and serves several essential functions.

In the adult skeleton, the first step in remodeling is almost always bone resorption; the old material has to be cleared away before new bone can be deposited. In reabsorbing bone, osteoclasts attach to a bony surface and secrete acid to dissolve the mineral and proteolytic enzymes to digest the matrix. They thereby physically remove a volume of bone. The calcium content of that volume is released into the bloodstream and becomes available both to support the level of calcium in the extracellular fluids of the body against the various drains to which it may be subject and to meet the calcium demands of bony sites elsewhere in the body that happen currently to be in their mineralization phase.

Bone is formed by osteoblasts, which first deposit a protein matrix on an excavated surface and then act on it to create crystal nuclei of calcium phosphate. Thereafter, these nuclei grow by accretion, without further cell work, spontaneously adding calcium and phosphate ions drawn out of the blood that flows past the mineralizing site. Once deposited at any given site, calcium is permanently trapped and can be removed only by the process of bone resorption at that site.

Remodeling fluxes into and out of bone, in the mature adult, are typically in the range of 300 to 500 milligrams of calcium (mg Ca) per day, or about two to four times as large as the aggregate of the external calcium fluxes (absorption and excretion).

Calcium Homeostasis

The balance between bone formation and resorption is adjusted so as to keep the calcium concentration of the extracellular fluid constant. The process is mediated mainly through the action of what are termed "calcitrophic hormones" – principally parathyroid hormone, calcitonin, and calcitriol – with parathyroid hormone being the most important in mature adults, and calcitonin and calcitriol notably more important in infancy.

Parathyroid hormone secretion is evoked by a fall in extracellular fluid calcium level, and calcitonin secretion by a rise. Parathyroid hormone acts to raise falling calcium levels by activating bone remodeling, by reducing renal calcium losses, and by increasing renal synthesis of calcitriol, the active hormonal form of vitamin D (which enhances intestinal absorption efficiency). Activation of remodeling helps raise a falling calcium level because resorption precedes formation, and thus in its early phases, remodeling provides a temporary surplus of calcium. Calcitonin, in contrast, lowers elevated calcium levels by temporarily suppressing bone resorption, thus stopping the release of calcium from bone.

The activity of maintaining the constancy of extracellular fluid calcium concentration is termed "calcium homeostasis." The body's ability to adjust bone remodeling balance is an important physiological defense of the calcium levels in the extracellular fluids of the body, providing needed calcium when the level would otherwise drop and soaking up surplus calcium when it would otherwise rise.

Given the calcium abundance in the diets of virtually all mammals, the reserve function in subhuman species operates mainly during periods of excessive skeletal demand or transient environmental scarcity. Such episodic withdrawal from the reserves is illustrated most clearly in what happens during antler formation in several species of deer each spring. Antlers consist of bone; their growth is usually so rapid that absorbed food calcium cannot keep up with demand, particularly given the relatively poor nutritional quality of early spring food sources. Accordingly, parathyroid hormone secretion increases sharply when antler formation begins, and a burst of bone remodeling is initiated throughout the skeleton. Because the initial phase of remodeling is resorptive, a temporary surplus of calcium is made available for antler mineralization. Later, antler growth slows or stops, and the remodeling loci throughout the skeleton enter their own phase of bone formation (which proceeds at a somewhat slower pace than for the antlers). Those sites then get the calcium they need from a diet that now contains calcium-rich summer grasses and foliage.

Averaged over the year, environmental calcium is usually quite sufficient to permit deer to build and to discard all that accumulated antler calcium annually, and then to start the process all over again. Remodel-

ing is adjusted in this case to help with a temporary calcium “cash-flow” problem.

So long as the microscopic scaffolding of bone from which calcium is borrowed remains intact, as in the deer, there is always the potential for restoration of most or all of the bone lost through remodeling imbalances. But this is only true if adequate exogenous calcium becomes available in time. This borrowing mechanism creates a structural problem for the skeleton when absorbed dietary calcium remains chronically below the demand created by daily losses. Since, under those circumstances, the calcium borrowed from bone cannot be repaid, the remodeling imbalance continues and bone mass continues to be eroded. If this process reaches the point where structural elements are lost (for example, trabecular plates are perforated or trabecular spicules disconnected), much of the loss becomes effectively irreversible, and the deficiency can no longer be corrected, at least by restoring the missing nutrient.

Intrinsic Control of Skeletal Mass

While the ability of the skeleton to release calcium for homeostatic purposes (by tearing down its own bony substance) is effectively limitless, the ability to store excess calcium is much more limited. This is because, as has already been noted, calcium can be stored only by forming new bone in excess of the amount resorbed. But there has to be some ceiling here. Otherwise, in a typically calcium-rich environment, higher vertebrates, storing continuing surpluses of calcium, would become all bone.

How much bone an organism possesses when calcium intake is not the limiting factor depends mainly upon the degree of mechanical strain each bone experiences. Throughout the terrestrial vertebrates each bone adjusts its density (through balancing resorption and formation) so that it experiences in the range of 1,000 to 1,500 microstrain in ordinary use. (Strain is the bending any structure undergoes when it is loaded; 1,000 microstrain is a dimensional deformation of 0.1 percent.) So far as is now known, no surplus of nutrients will lead to more bone accumulation than what is required to produce and maintain that degree of stiffness. Thus, homeostatically, bone mass is adjusted to support body fluid calcium levels; and structurally, bone mass is regulated to produce an optimal stiffness (not too massive, not too flimsy).

The control system regulating this structural aspect of bone mass is not fully understood, but it is known to be site-specific and to be intrinsic to bone (rather than extrinsic as with calcium homeostasis). What this system amounts to is that local bone formation exceeds local bone resorption when bone deforms excessively, making it stiffer, and the opposite occurs when local bone deformation is minimal. Thus, like muscle, bone hypertrophies with use and atrophies with disuse.

Both the intraosseous stiffness-optimizing system and the extraosseous calcium homeostatic system alter bone mass by regulating the balance between bone formation and bone resorption, that is, they both use the remodeling apparatus to alter bone density. In certain circumstances they reciprocally influence one another. For example, when the homeostatic system acts to reduce density, it thereby leads to increased strain on routine loading of the skeletal region involved. This, in turn, creates a signal to restore lost bone as soon as environmental calcium becomes available once again.

This departure from optimal mass levels is always downward – borrowing and then paying back. There is only limited capacity to store calcium above current structural needs for bone. Although the homeostatic surplus is literally vast, relative to metabolic functions of calcium, there is virtually no bodily ability to build a structural surplus, at least relative to current levels of mechanical usage. Instead, the structural reserve of the skeleton lies in the fact that normal bone can withstand greater deformation than the 1,000 to 1,500 microstrain of everyday use. The limit is closer to 7,000 microstrain, but its actual value depends upon how rapidly and how often a load is experienced. This margin of safety is what protects us from fracture when we experience low-level falls and bumps.

Osteoporosis

Definition and Expression

Osteoporosis is a disorder of bone characterized by excessive fragility due either to a decrease in bone mass or to microarchitectural deterioration of bone tissue (or both). It is a structural weakness in an organ system that, as has already been noted, serves as a source and a sink for calcium in its primary evolutionary function.

The bony fragility that constitutes osteoporosis is expressed in a propensity to develop fractures on minor injury. This fragility may involve virtually any bone in the skeleton. Stereotypical fracture syndromes involve such regions as the spine, the upper end of the femur (hip fracture), and various extremity sites, for example, wrist and shoulder. But ribs, pelvis, hands, and feet are also common fracture sites in patients with osteoporosis.

Bases for Bony Fragility

Osteoporosis is not a unitary disorder and does not have a single pathogenesis. Basic engineering considerations make it clear that the strength and stiffness of bone, as is true for any structure, derive from four main sources: the intrinsic physical properties of its component material; the mass density of that material; the spatial arrangement of the material; and the loading history of a given member (which expresses itself in an accumulation of ultra-

microscopic defects called “fatigue damage”). When any structure fails under load, it is because of relative weakness due to insufficiency of one or more of these strength determinants.

In most of the osteoporotic fracture syndromes there is, currently, no recognized abnormality of the bony material itself. The bony substance in the skeleton of patients with osteoporosis is qualitatively much like the bony substance in normal individuals. Instead, the principal bases for weakness in osteoporotic bone are to be found in (1) a decreased amount of bony material, or mass density (to which the term “osteoporosis” literally refers); (2) accumulated fatigue damage in that bone which is present; or (3) architectural defects in the latticework of trabecular (or cancellous) bone. These latticework defects, in turn, are of two types, microfractures of trabecular elements, which have previously occurred under loading but have not yet fully healed (and which thereby render the latticework weak), and preferential severance (and loss) of the horizontal bracing trabecular elements that give the lattice much of its stiffness.

Interactions of several of the more important contributing factors are illustrated in Figure IV.D.4.1. Where nutrition, and specifically calcium, come into this complex interplay of fragility factors is predominantly through their effect on bone mass density, that is, through the size of the calcium reserve. Thus, although important, calcium intake is only one of several interacting factors that can lead to osteoporosis. Some individuals will develop fragility fractures because bone mass is reduced, but others will develop them because of failure to repair fatigue damage or because of defective trabecular architecture. Even in regard to decreased bone mass, calcium shares the stage with other important factors such as gonadal hormone deficiency, physical inactivity, and a variety of lifestyle factors such as alcohol abuse and smoking. These factors also reduce bone mass, but their action on bone is largely independent of nutrition. (High blood alcohol levels poison bone cells, just as they do cells of other tissues. Hence, it is not

surprising that bone tissue fails variously in habitual alcohol abusers. The mechanism of the effects of tobacco is unknown. Smoking women have earlier menopause and lower postmenopausal estrogen levels than nonsmokers, but this is probably only part of the explanation.)

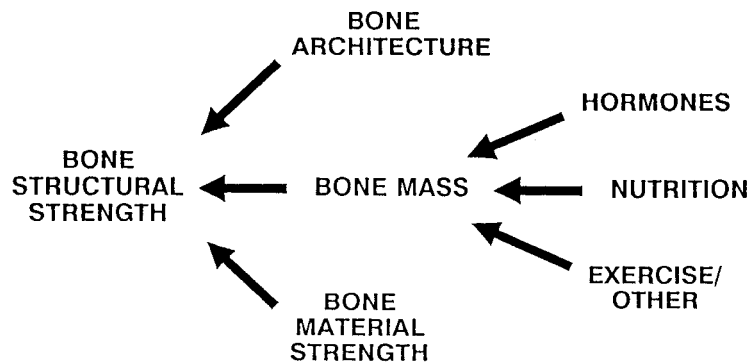
Until the multifactorial character of osteoporosis causation was fully understood, there had been confusion about the importance of calcium, mainly because published studies did not always show a protective effect of an adequate calcium intake. Figure IV.D.4.1 forcefully emphasizes why a universal protective effect is an unrealistic expectation. All that an adequate calcium intake can do is to help the organism build the largest possible skeleton during growth and to protect the skeleton against one-sided calcium withdrawals during maturity. But a high calcium intake will not counteract, for example, effects of alcohol abuse or physical inactivity.

Nevertheless, available evidence suggests that if a fully adequate calcium intake could be assured for every member of the population, as much as 50 percent of the osteoporosis burden of the developed nations would be alleviated. Even so, there would still be 50 percent that persists. These cases would have bases other than nutritional inadequacy.

Osteoporosis: A Disorder of a Nutrient Reserve

Although osteoporosis, when caused by inadequate calcium intake (either during growth or during the adult years), can be said to represent a nutritional disorder, it is important to recognize that the primary metabolic function of calcium is never even remotely compromised. Thus, those forms of osteoporosis that result from inadequate calcium intake can be said to be disorders of a nutritional reserve, and not nutritional deficiency in the usual sense (as might occur with vitamin C and scurvy, or with vitamin D and rickets). Fat may be the only analogue of this unique nutritional situation, serving not only as an energy reserve but as insulation for warm-blooded organisms living in a cold environment.

Figure IV.D.4.1. Causal connections of some of the major factors influencing bone strength. Nutrition is only one of the factors that influence bone mass, and bone mass, in turn, is only one of the factors influencing bone strength. Bone material strength, in this diagram, includes the effect of fatigue damage discussed in the text. (© Robert P. Heaney 1991; reproduced with permission.)



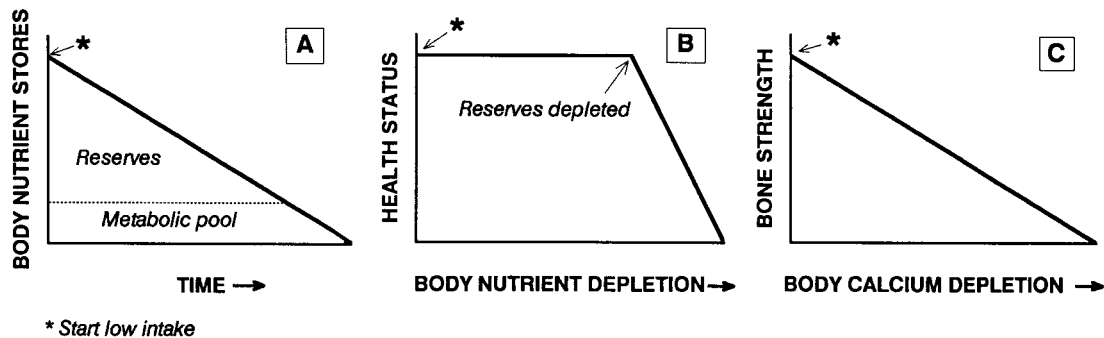


Figure IV.D.4.2. Schematic illustration of the relationship between body depletion of a nutrient and health status: (A) the depletion of body stores of a nutrient after placing the organism on a deficient intake; (B) the pattern common to most vitamins, in which health is maintained unchanged until the reserve is exhausted; (C) the pattern exhibited by calcium, in which health, as reflected in skeletal strength, declines steadily as the reserve itself is depleted. (© Robert P. Heany 1992; reproduced with permission.)

The difference between these two types of nutritional deficiency is illustrated schematically in Figure IV.D.4.2, which contrasts the effects produced by depletion of the body content of a nutrient such as vitamin C with the effects of calcium depletion. In the former case, where the reserves are only that, and serve no other function, health is maintained until the entire reserve is depleted. Then, as the active metabolic pool declines, dysfunction develops. With calcium, by contrast, any depletion of the reserve produces a corresponding decrease in skeletal strength. The skeleton would be rendered totally useless as a structure long before the metabolic pool of calcium would be compromised.

One of the problems this arrangement creates for the nutritional scientist is that a deficiency of calcium sufficient to compromise the structural dimension of the reserves will have no impact upon the basic metabolic function of calcium. Thus, a calcium deficiency sufficient to produce osteoporosis will not be reflected in appreciable decreases in calcium concentration in the circulating fluids, nor in the critical cell compartments where calcium functions as a second messenger, nor even in the ready availability of calcium ions for that crucial function. For this reason there are no blood or urine tests that, alone or in combination, are diagnostic either for this phase of calcium deficiency or for osteoporosis.

Calcium Requirement

Definition

A requirement for a specific nutrient has traditionally been defined as the intake necessary to prevent the expression of the disease or disability associated with deficiency of that nutrient. In recent years

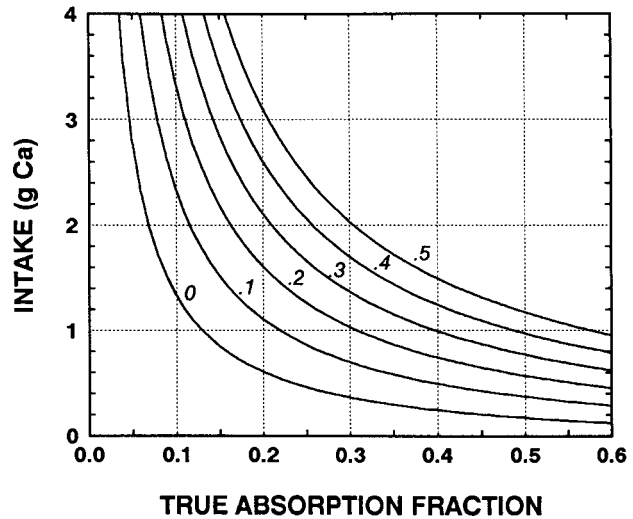
there has been a tendency to broaden that definition to read: the intake required to promote optimal health. In the case of calcium, neither approach is particularly apt, as the foregoing discussion has emphasized, inasmuch as health in this instance is not a matter of the basic metabolic function of calcium but of the size of the calcium reserve. Hence a calcium requirement relative to bone health needs to be defined as the intake necessary (1) for building the largest bone mass possible within the genetic program of the individual, and (2) for maintaining that reserve against unbalanced withdrawals after growth has ceased. Any unrepaired decrease in the size of the reserve, other things being equal, reduces bone strength.

The Bases for a Calcium Requirement

Because providing during times of need is precisely what a reserve is for, this process cannot properly be considered harmful in itself. A problem develops only when the process is one-sided, with reserves drawn down but not replenished. In this connection, it will be useful to review here both certain quantitative aspects of how the body maintains constant calcium levels in its blood and extracellular fluids, and how and why an unbalanced situation develops.

As mentioned previously, foods available to high primates and hunter-gatherer humans in their natural states are rich in calcium. In fact, nutrient densities of such foods average in the range of 70 to 100 mg Ca/100 kilocalories (kcal). When calculated for an energy intake sufficient to sustain a hunter-gatherer of typical body size, this density range translates to a daily calcium intake of 2,000 to 3,000 mg, substantially in excess of what most modern humans get, and four to six times what an adult female in the United States typically ingests.

Figure IV.D.4.3. Relationship of calcium intake, absorption efficiency, and net absorption. All the points along a single curve represent the same net absorption value and thus define, for that value, what the intake would have to be to achieve that level of net absorption for any given absorption fraction. (© Robert P. Heaney 1991; reproduced with permission.)



Only after the advent of cultivated cereal grains at the agricultural revolution did humans shift from a diet with a high calcium density to one with a low calcium density.

Intestinal absorption. As already noted, calcium absorption is inefficient, averaging in the range of 30 percent of ingested intake in healthy adults at intakes in the range of the current Recommended Dietary Allowance (RDA) (800 mg for adults). Absorption efficiency varies inversely as roughly the square root of ingested intake, which means that although efficiency rises as intake falls, the evoked rise in absorption fraction will not be sufficient to offset the actual drop in intake.

Furthermore, digestive secretions themselves contain calcium (typically in the range of 150 mg/day in a healthy adult). Because this calcium is secreted along the length of the gut, it is reabsorbed at only about 85 percent the efficiency of ingested calcium (which is itself absorbed inefficiently). This secreted calcium constitutes a cost of digestion, and most of it will be lost in the feces.

This two-directional traffic means that net absorption will always be less than gross absorption. For example, at an intake of 600 mg/day and an absorption efficiency of 30 percent, net absorption will be only about 63 mg, or barely 10 percent of intake. In fact, at low intakes, net absorption will commonly be negative, that is, there will be more calcium going out in the feces than is coming in by way of the mouth. This does not represent intestinal pathology but is an inevitable consequence of the combination of low absorption efficiency and calcium secretion with the digestive juices.

The quantitative character of this relationship is depicted schematically in Figure IV.D.4.3, which shows various iso-absorption contours relating intake and absorption efficiency to various values for net absorption (from 0 to 500 mg Ca/day). For example,

to achieve a net absorption of 200 mg/day (close to the figure required to offset extraintestinal losses in healthy adults), intake must be 1,030 mg at an absorption efficiency of 30 percent, and 1,605 mg at an absorptive efficiency of 20 percent.

The fact of low, and even negative, net absorption might be construed to indicate that the organism does not need much calcium after all, but that would be a misreading of the evidence. With the naturally high calcium intake of the hunter-gatherer diet, the customary exposure of the gut is to a calcium-rich ingestate, and even with low absorption efficiency, net absorption would always be positive. But in fact, the low absorptive performance of the gut in mammals is an evolutionary adaptation to this environmental surfeit. Unfortunately, a low absorption fraction is maladaptive for modern diets. Our bodies have adapted to a high-calcium diet, and the time span from the agricultural revolution to the present has been much too short to have allowed evolution toward greater absorptive efficiency.

Some nutrient factors interfere with calcium absorption and thereby raise the calcium requirement. Certain kinds of fiber bind calcium and thereby reduce its absorption. Wheat bran does this, for example. However, this is not true for all fiber: The fiber of green leafy vegetables does not interfere at all with absorption. Overall, the effect of fiber in our diets is relatively small, and even widely ranging fiber intakes would not be expected to exert very large effects on calcium absorption. A second factor reducing calcium absorption is caffeine, a substance widely considered among the lay and the scientific communities to be a risk factor for osteoporosis. It turns out, however, that the caffeine effect is also small and, except for very high daily coffee intakes, would not pose much of a problem. For example, the negative effect on calcium absorption of a single brewed cup of coffee is such

that calcium balance deteriorates by about 3 mg. This quantity is so small that its impact can be easily offset by as little as an ounce of milk.

Renal and dermal losses. Urinary and dermal calcium losses are also important determinants of nutrient requirement. Dermal losses occur through sweat and also through shed epithelial structures (dry skin, nails, and hair – all of which contain calcium). For the most part dermal losses are unregulated and thus constitute an irreducible daily drain that must be offset from the diet or, failing that, from the bony reserves.

Urine calcium in humans is determined predominantly by body size and by intake of protein and sodium and only to a lesser degree by calcium intake itself. Protein intake increases urinary calcium loss largely through the catabolism of sulfur-containing amino acids and the excretion of the consequent sulfate load (an endogenous equivalent, as it were, of the acid rain problem produced by burning sulfur-containing fossil fuels). High sodium intake also tends to wash calcium out through the kidneys. High sodium intake is another major dietary change that has occurred for the most part only very recently. Hunter-gatherers and high primates typically have sodium intakes nearly two orders of magnitude (something on the order of 100 times) lower than modern-day humans. In Switzerland, where careful records of salt imports have been kept, per capita salt consumption has increased by a factor of 10 since 1850 alone.

Urine calcium rises with body weight from infancy to adolescence. Infants can reduce urinary calcium losses to near zero when calcium intake drops. Children have an intermediate ability to conserve calcium at the kidney, and adolescents, strikingly, maintain high urinary calcium values irrespective of their calcium intake. This makes them, at a time when calcium requirements for bone growth are at their absolute maximum, particularly vulnerable to environmental scarcity. A partial explanation for the change with age may lie in the fact that infants and small children have relatively low salt intakes, and instead of burning ingested protein as fuel, they build at least some of it into new tissue. Thus neither of the factors driving the higher urine calcium values of older persons are prominent early in life.

The combination of unrecovered digestive juice calcium, minimal urinary calcium loss, and dermal loss, which occur under conditions of calcium restriction, constitutes what is called “obligatory loss.” In normal adults consuming typical Western diets, this loss amounts to something in the range of 150 to 250 mg/day, probably more often on the higher side of the range than on the lower. Whenever the absorbed intake is insufficient to compensate for this obligatory loss, bone remodeling will be adjusted so that resorption exceeds formation and thus bone mass will decrease. In other words, the organism calls upon its calcium reserves.

Recommended Dietary Allowances

A Recommended Dietary Allowance (RDA) is a population-level estimate of the daily intake that would be sufficient to meet the actual requirement for a given nutrient for virtually every member of the population. RDAs are deliberately set above the average requirement so as to assure meeting the needs of those with above-average needs.

The RDAs for calcium in the United States are 800 mg/day for children up to age 11, 1,200 mg/day from ages 12 to 24, and 800 mg/day thereafter. (During pregnancy and lactation, the RDA rises to 1,200 mg/day.) There is growing evidence that these requirements are low, and what follows will summarize the best current information about what the true requirement may be. In this connection it is worth recalling once again that wild plant foods are rich in calcium. As a result, the human calcium intake for foragers was almost certainly substantially above not only current intakes but even the RDAs. While there has been a bias in the scientific community in favor of current dietary practices, there should be no surprise when evidence indicates that intakes closer to (though still on the low side of) what our foraging forebears apparently got, may in fact better meet our needs.

Requirements during Growth

Threshold intake. One aspect of the relationship of calcium to bone health, relevant to a consideration of requirement, is the fact that calcium is a threshold nutrient. This means that calcium intake will limit the amount of bone a growing organism can acquire, at least up to some threshold value, and that above that intake further increases will no longer have any effect on bone mass accumulation. This concept is illustrated in Figure IV.D.4.4, which shows, first in Panel A, what the idealized relationship between intake and bone mass accumulation during growth would look like, and then, in Panel B, what experiments studying the effect of calcium intake on bone accumulation during growth have actually found in laboratory animals.

As far as bone is concerned, the optimal intake during growth would be an intake at or above the threshold, and the RDA would be a value sufficiently above the *average* threshold value to allow for the fact that different individuals will have differing values at which the threshold occurs. This interindividual variation reflects differences in absorption efficiency and in ability to restrict urinary loss.

It is important to recognize, in this regard, that calcium intakes below the threshold value will not necessarily limit growth in stature. It takes very serious depletion for that effect to occur, and even then it is hard to be certain that a low intake of calcium is the responsible factor, since intake of other essential nutrients is commonly reduced under such circumstances, as well.

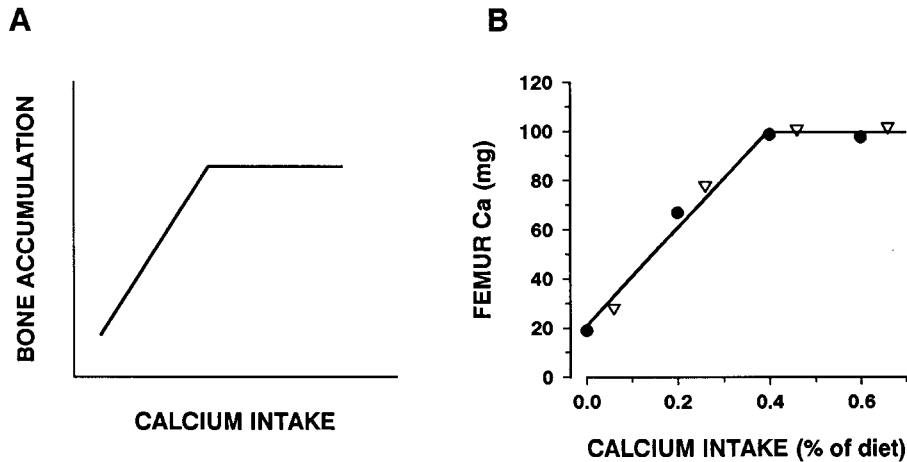


Figure IV.D.4.4. Threshold behavior of calcium intake: (A) Theoretical relationship of bone accumulation to intake. Below a certain value – the “threshold” – bone accumulation is a linear function of intake; in other words, the amount of bone that can be accumulated is limited by the amount of calcium ingested. Above the threshold, bone accumulation is limited by other factors and is no longer related to changes in calcium intake. (B) Actual data from two experiments in growing rats, showing how bone accumulation does, in fact, exhibit a threshold pattern. (Redrawn from data in Forbes, Weingartner, Parker, et al. 1979, in *Journal of Nutrition* 109: 1652–60; © Robert P. Heany 1992; reproduced with permission.)

What happens during growth under conditions of suboptimal calcium intake is, instead, simply that the skeleton achieves its genetically programmed external size and shape but is internally flimsier, meaning the bone cortices are thinner and more porous, and the trabeculae thinner and more widely spaced. Less mass is spread over a growing volume and the structure is thus intrinsically weaker.

While it is relatively easy to do threshold experiments in laboratory animals (as in Figure IV.D.4.4), it is much harder to do (or to justify) such experiments in growing children. Instead of the threshold approach, much of the judgment in regard to calcium requirement and RDAs during growth has been based upon actual food practices in populations that seem to be developing “normally.” There is an obvious circularity in that reasoning, because if current bone mass values are normative, then current intakes are manifestly sufficient to support “normal” bone development. Thus current intakes conform with the requirement, the average child is getting what he or she needs, and all is right in this best of all possible worlds.

However, osteoporosis has now reached epidemic proportions in the developed nations, and failure to develop the peak bone mass genetically programmed has to be considered a partial explanation. Thus current intakes, although “normal” in the sense of being usual, are not necessarily optimal.

Estimating the threshold intake. A traditional approach to the determination of nutrient requirements has been to perform what is technically referred to as a metabolic balance study – a study in which human subjects live in a laboratory environment and in which intake and output of the nutrient under study are carefully measured. The balance between intake and output is computed, and for nutrients that are not altered by their metabolism, such as minerals like calcium, that balance is a surrogate for body retention or loss, that is, skeleton building or destruction.

One of the problems with the metabolic balance technique is that although theoretically ideal, the result that it produces is inherently imprecise. That problem can be minimized by doing a large number of such studies, because the uncertainty range about the average estimate of balance declines as the square root of the number of studies. But balance studies are so difficult and expensive to perform that few investigators can accumulate a large enough experience to give the required precision at the several levels of intake that must be studied in order to determine the requirement.

Velimir Matkovic of Ohio State University has attempted to solve this problem by assembling all of the published balance studies performed on growing humans, spanning a 70-year period from 1922 to 1991. After excluding those that did not meet certain

a priori criteria, he was able to assemble over 500 published studies, a number large enough to give some reasonable certainty to the estimates and also to ascertain whether the threshold behavior observed in animals (Figure IV.D.4.4) is also found in growing humans.

He and I, working together on this project, found that the threshold behavior did, in fact, apply to growing humans, one example of which is shown in Figure IV.D.4.5. This work also allowed us to estimate the average threshold intake at various stages during growth. These threshold values are presented in Table IV.D.4.1, by age group from infancy through age 30. Equivalently, these values are average requirements for the population of growing humans. As already noted, the corresponding values for RDAs would be higher still. It can be seen that all of the threshold values in Table IV.D.4.1 are above the current RDAs in the United States. Nevertheless, intakes at or above these threshold values would have been common in human hunter-gatherers living in equilibrium with their environment, and hence, although such intakes would be atypical by modern standards, they can hardly be considered unnaturally high.

Corroborating evidence. This means of estimating requirements during growth is probably the best available approach to the problem, and although it produces values above the current RDAs, it is important to recognize that its estimates are not, in fact, at variance with other recent data, which tend to show that, other things being equal, more calcium intake during growth leads to more bone mass accumulation. Thus, very recently C. C. Johnston, Jr., and his colleagues from Indiana University reported results of a double-blind, placebo-controlled study in identical twin children. Calcium supplementation given to one member of a twin pair produced more bone mass accumulation than in the unsupplemented twin.

A striking feature of that study was the fact that the unsupplemented members of each twin pair averaged an intake that was already above the official RDA for calcium for the age concerned. This behavior suggests that the official RDA value is below the threshold (see Figure IV.D.4.4).

Similarly, R. R. Recker and his colleagues from Creighton University have recently reported results from a longitudinal study of bone mass accumulation in women aged 20 to 30. They showed both that bone mass continues to accumulate at a rate of about 1 percent per year to age 30, and that calcium intake (specifically the calcium-to-protein ratio of the diet) was the single most important determinant of how much bone was added during that decade.

Requirement during Maturity

From age 30 until menopause in women, and from age 30 until the beginning of senescence in men, the requirement for calcium can be defined as the intake necessary to offset obligatory losses, or, put another way, the intake needed simply to keep the body in equilibrium, neither gaining nor losing calcium. Current estimates indicate that the mean requirement is in the range of 800 to 1,200 mg/day for typical Western diets for both men and women. It would probably be substantially less in individuals who have low protein or low sodium intakes (or both). This is because, as already noted, both high protein and high sodium intakes increase urinary calcium loss and hence decrease the ability of the body to conserve calcium when calcium intake is reduced.

Incidentally, this observation is not to suggest that calcium is somehow a “good” nutrient and that protein and sodium are somehow “bad.” Rather, it simply emphasizes that requirements are not abstract absolute values but are reflections both of metabolic activity of the organism and of other constituents in the diet.

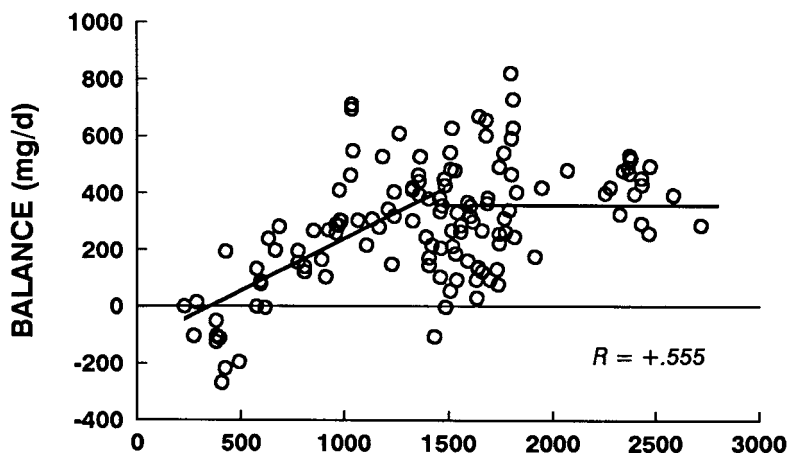


Figure IV.D.4.5. Relationship of calcium intake to calcium balance in adolescents. Above the threshold intake value (1,480 mg/day), balance no longer changes with intake, whereas below the threshold, balance and intake are directly correlated. (© Robert P. Heaney 1991; reproduced with permission.)

Table IV.D.4.1. *Threshold calcium intakes during growth*

Age (in years)	Threshold intake (mg/day)
0-1	1,090
2-9	1,390
9-17	1,480
18-30	957

If one can judge from the food habits of contemporary hunter-gatherers, they would not have had much experience with sodium; however, high-protein diets would have been common. Meat is a very efficient food source, rich in many essential nutrients, and some studies have suggested that for human foragers, protein intake might have accounted for as much as 35 percent of total calories – a higher intake than even that achieved by typical citizens of developed nations. But the diet of hunter-gatherers, as has already been noted, was also very high in calcium, and so a high protein intake would not have created an effective calcium deficiency as it does when calcium intake is as low as is commonly found in U.S. diets.

Requirements during Postmenopause and Senescence

During the declining years, the calcium requirement can no longer be defined as the intake required to offset obligatory losses. This is because bone loss occurs now for intrinsic reasons as well as for homeostatic ones. Even the richest of diets cannot totally prevent this kind of bone loss.

Menopausal bone loss. The first example is the bone loss that occurs in women during the early postmenopause (from cessation of menses to 5–10 years later). To the extent that there has been any controversy or apparent disagreement about the importance of an adequate calcium intake in adults, that controversy has centered around what occurs during this brief period in a woman's life. This is a time when typical women lose something approaching 15 percent of the bone mass they possessed immediately before cessation of ovarian function. Virtually without exception, studies of calcium intake or calcium supplementation during this time have shown that calcium has little or no effect on this bone loss.

There is now general agreement that bone loss at this time is due almost exclusively to loss of gonadal hormones. (The same type of loss follows castration in males.) While many of the details of the mechanism remain uncertain, it can be said that this loss reflects a downward adjustment of bone mass to a new steady state, just as occurs with immobilization. During that approach to a new postmenopausal equilibrium value, nonhormonal forces, such as calcium or exercise, are without effect – simply because the change in bone mass is due specifically to the diminution of

gonadal hormones and not to low calcium intake or inadequate exercise. In fact, during the few years after menopause, so much calcium may be made available from bone that there may be no external calcium requirement at all. However, when the new steady state is approached and bone loss slows, all the old interactions reappear.

This conceptual framework was not available to investigators until recently, which may explain why so many previously published studies of the effect of calcium chose to address the early postmenopausal period. Those years are, of course, the time when bone loss is the most rapid and the value of successful intervention is most evident, so it is not surprising that it has been extensively studied. But loss at that life stage is caused by estrogen lack and is best prevented by estrogen replacement.

Senescence. Bone loss during senescence in both men and women is a complex process and has many determinants. This is to some extent true earlier in life as well, but nonnutritional factors loom larger during the declining years of life. As a general rule, mechanical loading on the skeleton decreases with age, in part because older people do less strenuous work and in part, also, because with maturity they become more graceful and efficient in what they do. Thus, some downward adjustment in bone mass simply reflects a decline in mechanical need. Inevitably that decreases the effective strength reserve, which is useful to resist the impact of falls. As already noted, the mechanical adjustment system responds to current usage, not to potential injury. This type of decline in skeletal mass is not nutritionally related and, in the practical order, cannot be appreciably altered by assuring a high calcium intake.

However, a number of changes also occur in the calcium economy of the aging person. There is a decline in absorption efficiency at the intestine, a partial resistance to vitamin D action on the intestine, and a decrease in the ability to synthesize calcitriol, the active hormonal form of vitamin D. Additionally, in women who are deprived of estrogen after menopause, there is deterioration in the ability to conserve calcium at the kidney. For all of these reasons, older persons absorb calcium less well from the diet and retain less of what they do absorb.

This means that calcium intake requirement rises in the elderly. Unfortunately, actual intake tends to go in the wrong direction. With a decline in physical activity, there is a tendency for food intake itself to go down, and that almost always reduces the intake of all nutrients, calcium among them. This combination of increased need and decreased intake in the elderly sets the stage for intake-related bone loss.

The majority of investigations of the relationship of calcium intake to bone status in the elderly – particularly the double-blind, placebo-controlled trials of B. Dawson-Hughes and her colleagues at Tufts Univer-

sity and of Petra Elders and her colleagues at the Free University of Amsterdam – show that additional calcium intake will reduce, to a substantial extent, the degree of bone loss that is otherwise occurring in ostensibly normal, healthy individuals. Total intakes shown to be adequate for this purpose have generally been in the range of 1,000 to 2,000 mg of Ca/day. As already noted, although such intakes are higher than current practice, they cannot be considered high when compared with the typical intakes of our foraging ancestors.

The fact that published studies show such a dramatic reduction in what had otherwise been considered an inevitable, age-related loss of bone indicates that a substantial fraction of that loss was not inevitable after all but was diet-related. This conclusion is even more forcefully emphasized by the findings of three important European studies. One, performed in France, randomized more than 3,200 elderly women (average age 84) to receive either a placebo or 1,200 mg calcium plus 800 IU vitamin D. Not only did the supplemented women stop losing bone, but fractures also declined dramatically within 18 months of starting treatment. The other studies, one undertaken in Finland and the other in Switzerland, although differing somewhat in design from the French study, nevertheless also clearly showed fracture reduction in elderly individuals given calcium or vitamin D supplements.

Miscellaneous Considerations

Calcium Intake and Bone Remodeling

We have already noted that one of the causes of bone fragility is the accumulation of unrepaired microscopic fatigue damage, something that occurs inevitably in all structural materials when they are loaded, but which, given the living character of bone in most vertebrates, is susceptible of repair. That repair is accomplished by bone remodeling, which has been visited, so far in this chapter, mainly as a means of altering bone mass. In fact, bone remodeling has the dual function of both adjusting mass and replacing damaged bone. Failure to effect that replacement in a timely fashion permits fatigue damage to accumulate to the point where serious structural failure can occur on application of relatively small forces.

Effecting this repair depends first upon the body's ability to sense the presence of microscopic damage. Available evidence suggests that the threshold of bone sensitivity to such damage is determined to a substantial extent by the circulating level of parathyroid hormone (PTH). But PTH secretion, in turn, is influenced predominantly by calcium need, and ultimately, therefore, by calcium intake. Thus, a constant high-calcium diet, particularly with the intake spread out over the course of the day, leads to low PTH levels, and therefore to a low sensitivity of the apparatus

that detects fatigue damage. The practical import of these considerations is not known for certain, but several investigators have expressed concern about the possible dangers of a constant suppression of the remodeling process.

How could this be a problem if, as already noted, human hunter-gatherers had calcium intakes substantially higher than we now experience? Presumably, their PTH levels would have been substantially below ours. However, there is no fossil evidence that hunter-gatherers suffered an undue fracture burden. Thus either the PTH levels in human foragers were not, in fact, constantly low, or the threshold for detection of fatigue damage is below even the low PTH levels produced by a chronic high calcium intake. The former seems to be the more likely explanation.

Although *average* calcium intakes would have been high for the hunter-gatherers, it becomes clear on reflection that under field conditions, food intake, and thus calcium intake, could not have been a constant affair. There would have been inevitable periods of fasting and of environmental scarcity. In fact, the PTH mechanism evolved precisely to handle such times of calcium need, and its very presence in mammals may be taken as presumptive evidence of at least periodic calcium deprivation.

In addition to seasonal variation in food availability, females were subject to the regular predictable calcium drain of pregnancy and lactation. Another example of periodic need, already cited, is antler formation in deer. At the beginning of antler formation, there is a burst of PTH-mediated remodeling throughout the skeleton. This causes, as it were, a flurry of "spring housecleaning," which, if it had not occurred earlier, serves to remodel areas of fatigue damage that had accumulated in the skeleton over the preceding year. In this way, antler formation in male deer produces a drain analogous to that produced by pregnancy and lactation in female mammals generally.

Thus, even superficial reflection serves to make clear that the "natural" situation would have been one in which PTH is periodically evoked, and therefore one in which conditions would periodically be right for resorption and replacement of damaged bone.

That, however, is not the situation that obtains in civilized, adult, affluent humans who have few children, who lactate even less often, and who rarely experience prolonged periods of fasting or deprivation – in other words, typical adults in developed nations. Is a constant high intake of calcium optimal for these individuals? One cannot say with any assurance one way or the other. Possibly a periodic calcium "vacation" might be salutary. Possibly the once common Christian practice of fasting during Lent and four times a year at the Ember Days, and the still prevalent practice among observant Muslims of fasting during Ramadan, may evoke effects that are as salutary for the body as for the soul.

Recovery after Illness, Injury, and Disability

One common, if not quite universal, feature of even an affluent, protected, developed-world life is the fact of injury, illness, and disability, events that, however temporary and repairable, nevertheless enter the lives of most humans. Generally, these events are periods of reduced nutrient intake, sometimes of enhanced excretory loss of critical nutrients, and often periods of reduced mobility and physical activity as well. Reduced physical activity always leads to a reduction in bone mass, even in well individuals; such disuse loss is aggravated in the presence of illness.

What happens during recovery from such episodes? Usually, as far as a sense of well-being, health, and vigor are concerned, we return to our former status. But if our regular calcium intake is just sufficient to maintain equilibrium, that is, to offset daily losses, then, ipso facto, that intake will not be sufficient to permit replacement of the bone lost during the preceding illness or disability. We have no consciousness of our bone mass, so we are unaware that this part of our bodies has not fully recovered.

Periodic episodes of this sort over a person's adult life, and particularly during the declining years, can contribute significantly to age-related bone loss. Such loss, however, is not necessarily irreversible until, as already noted, it proceeds to the point where some of the bony scaffolding is lost. (Then there is no chance of rebuilding what was once there.) It would seem important, therefore, to ensure, during recovery from periodic episodes of illness or disability, that our calcium intake is augmented, for the simple and obvious reason that the intake must be sufficient, at that time, not only to offset obligatory losses but also to replace bone mass lost during the preceding weeks.

Calcium in the Treatment of Osteoporosis

Although calcium is a necessary component of most treatment regimens for established osteoporosis, it is rarely sufficient by itself. Calcium is *necessary* because it is not possible to restore lost bone without providing adequate quantities of the building blocks of bone. That means a high calcium intake. But calcium alone is rarely sufficient because a high calcium intake suppresses bone remodeling. Although this suppression helps to stabilize bone mass and slow bone loss, at the same time it impedes substantial gain in bone mass. Bone gain requires additional stimulation by osteogenic factors or medications (such as growth hormone or sodium fluoride). Then, with that kind of bone-building stimulus, an adequate calcium intake helps to assure that the weakened bone can be rebuilt without having to shift calcium from other skeletal regions – robbing Peter to pay Paul, as it were.

One might wonder why a high calcium intake is compatible with bone gain during growth yet does not act the same way in older persons with established osteoporosis. The answer is simply that during

growth, the body produces large quantities of tissue-building factors such as growth hormone. These factors are nearly absent in older persons with osteoporosis. Furthermore, the violent mechanical loading of the skeleton typical of childhood and adolescence creates a powerful local stimulus to strengthen the skeleton. By contrast, a partially disabled, hurting, elderly person with one or more osteoporotic fractures generally decreases his or her physical activity – a situation that leads to bone loss irrespective of calcium intake.

Because of lowered calcium absorptive efficiency both with age and with decreased physical activity, generally large quantities of calcium are required as part of an osteoporosis treatment regimen. The exact quantities are uncertain, but they probably fall in the range of 1,500 to 2,500 mg/day. This is more than most persons with restricted activity can get from readily available foods, and some resort will usually have to be made to calcium-fortified foods or to calcium supplement tablets.

Ethnicity

Although bony fragility is not confined to Caucasians, it is much more common among whites than among blacks or Asians, and among Caucasians it is more common among those of northern European ancestry. Partly for these reasons, most of the research relating calcium to osteoporosis has been done in Caucasians. Despite a generally lower calcium intake than their white counterparts, blacks in the United States have heavier bone mass at all ages, from infancy onward. Moreover blacks, both in Africa and in the United States, have lower fracture rates than whites, even after correcting for differences in bone mass. Asians have bone mass values at least as low as Caucasians and seem to lose bone with age much as do whites, but available data suggest that fragility fractures, particularly hip fractures, are less common among them. These ethnic differences in fracture rate probably reflect the fact, noted earlier, that fragility has many bases, and that it takes more than just bone loss to result in a fracture. For example, the angled upper segment of the femur is shorter in Asians than in Caucasians. Engineering analysis has shown that long upper segments are structurally weaker than short ones. This difference, based ultimately in genetics, has a mechanical, structural basis, rather than a nutritional one.

American blacks have been studied somewhat more intensively than other non-Caucasian groups, and thus more is known about their calcium economy. Norman Bell and his colleagues at the Medical University of South Carolina have shown that at least a part of the explanation for their greater bone mass is a generally higher resistance of their bones to serve as a calcium reservoir. Thus, in maintaining extracellular fluid calcium levels, blacks require a higher level of PTH to release calcium from bone than do whites.

As noted earlier, PTH also enhances both intestinal absorption and renal conservation of calcium. Thus the higher PTH levels of blacks result in better utilization of dietary calcium, with a consequent skeletal-sparing effect.

Conclusion

Calcium is a nutrient, not a drug. The only disorder it can be expected to prevent or alleviate is calcium deficiency. The skeleton serves both as structural support for our bodies and as a reserve of calcium needed to maintain the constancy of body fluid calcium levels. Calcium deficiency, as it affects us, is a reduction in the size of the reserve and thus in its structural strength. An adequate intake will ensure that calcium deficiency will neither cause skeletal weakness in its own right nor aggravate the weakness produced by other causal factors. But calcium will not prevent or reverse the bone loss and fragility due to other factors. The evidence indicates that calcium deficiency is prevalent in the adult populations of Europe and North America and that it contributes in an important way to their osteoporotic fracture burdens.

Robert P. Heaney

Bibliography

- Christiansen, C., ed. 1991. Proceedings of a symposium: Consensus development conference on osteoporosis. *American Journal of Medicine* 91: S5B.
- Dawson-Hughes, B., G. E. Dallal, E. A. Kralli, et al. 1990. A controlled trial of the effect of calcium supplementation on bone density in postmenopausal women. *New England Journal of Medicine* 323: 878–83.
- Eaton, S. B., and D. A. Nelson. 1991. Calcium in evolutionary perspective. *American Journal of Clinical Nutrition* 54: 281–7.
- Elders, P. J. M., J. C. Netelenbos, P. Lips, et al. 1991. Calcium supplementation reduces vertebral bone loss in perimenopausal women: A controlled trial in 248 women between 46 and 55 years of age. *Journal of Clinical Endocrinology and Metabolism* 73: 533–40.
- Forbes, R. M., K. E. Weingartner, H. M. Parker, et al. 1979. Bioavailability to rats of zinc, magnesium, and calcium in casein-, egg-, and soy protein-containing diets. *Journal of Nutrition* 109: 1652–60.
- Heaney, R. P. 1989. Osteoporotic fracture space: An hypothesis. *Bone and Mineral* 6: 1–13.
1990. Estrogen-calcium interactions in the postmenopause: A quantitative description. *Bone and Mineral* 11: 67–84.
1992. Calcium intake and bone health in the adult. In *Clinics in applied nutrition*, ed. R. Chernoff and R. P. Heaney, 10–29. Reading, Mass.
- Holbrook, T. L., E. Barrett-Connor, and D. L. Wingard. 1988. Dietary calcium and risk of hip fracture: 14-year prospective population study. *Lancet* 2: 1046–9.
- Johnston, C. C., Jr., J. Z. Miller, C. W. Slemenda, et al. 1992. Calcium supplementation and increases in bone mineral density in children. *New England Journal of Medicine* 327: 87.
- Matkovic, V., and R. P. Heaney. 1992. Calcium balance during human growth. Evidence for threshold behavior. *American Journal of Clinical Nutrition* 55: 992–6.
- Matkovic, V., K. Kostial, I. Simonovic, et al. 1979. Bone status and fracture rates in two regions of Yugoslavia. *American Journal of Clinical Nutrition* 32: 540–9.
- McCarron, D. A., M. Lipkin, R. S. Rivlin, and R. P. Heaney. 1990. Dietary calcium and chronic diseases. *Medical Hypotheses* 31: 265–73.
- Recker, R. R., K. M. Davies, S. M. Hinders, et al. 1992. Bone gain in young adult women. *Journal of the American Medical Association* 268: 2403–8.
- Recommended Dietary Allowances*. 1989. Tenth edition. Washington, D.C.

IV.D.5 🦋 Pellagra

Pellagra is a chronic disease that can affect men, women, and – very rarely – children. The onset is insidious. At first, the afflicted experience malaise but have no definite symptoms. This is followed by the occurrence of a dermatitis on parts of the body exposed to sunlight. A diagnosis of pellagra is strongly indicated when the dermatitis appears around the neck and progresses from redness at the onset to a later thickening and hyperpigmentation of the skin in affected areas. The dermatitis appears during periods of the year when sun exposure is greatest. Other symptoms, including soreness of the mouth, nausea, and diarrhea, begin either concurrently with the skin changes or shortly thereafter. Diarrhea is associated with impaired nutrient absorption, and as a result of both dietary inadequacies and malabsorption, pellagrins frequently show clinical signs of multiple nutritional deficiencies.

Late signs of pellagra include mental confusion, delusions of sin, depression, and a suicidal tendency. Occasionally, these psychiatric signs are accompanied by a partial paralysis of the lower limbs. In the final stages of the disease, wasting becomes extreme as a result of both a refusal to eat (because of nausea) and pain on swallowing (because of fat malabsorption). Death is from extreme protein-energy malnutrition, or from a secondary infection such as tuberculosis, or from suicide (Roe 1991).

Pellagra as a Deficiency Disease

Since 1937, it has been known that pellagra is the result of a deficiency of the B vitamin niacin (Sydenstricker et al. 1938), and that the deficiency usually arises as a consequence of long-term subsistence on a diet lacking in animal protein or other foods that would meet the body's requirement for niacin (Carpenter and Lewin 1985). However, a sufficient quantity in the diet of the amino acid tryptophan – identi-

fied as a “precursor” of niacin (meaning that the body can convert tryptophan into niacin) – has also been found to cure or prevent the disease (Goldsmith et al. 1952). This explains why milk, for example, helps combat pellagra: Milk contains relatively little niacin but is a good source of tryptophan. Pellagra is most strongly associated with diets based on staple cereals, especially maize. This grain has historically been the daily fare of those who develop the disease: Maize is high in niacin content, but much of this niacin is in a chemically bound form that prevents absorption of the vitamin by the body (Goldsmith 1956).

History

Early Observations in Europe

Pellagra was initially described during the first third of the eighteenth century, when Gaspar Casal, a Spanish physician, wrote of an illness that had arisen among the peasants of the town of Oviedo, who subsisted largely on maize. The disease was termed *mal de la rosa* because of a peculiar sunburn-like dermatitis – its telltale mark. Casal’s book, published posthumously (1762), includes a frontispiece showing a classical figure with a “rose” on the top of each hand and foot and a chain of “roses” around the neck. The “roses” appear as if branded on the skin, suggesting that their appearance betokened a stigmatized condition (Roe 1973).

Casal studied the multisymptom disease that appeared to affect only the poorest laborers in this sharecropping region. He described how the afflicted lived on a diet of maize flour made into flat, baked cakes. The only other items of their diet were a few turnips, chestnuts, cabbages, beans, and apples. Casal found that the disease could be treated by providing milk, cheese, and other foods of animal origin to those afflicted.

Shortly after pellagra made its appearance in northern Spain, it was seen in Italy, where there was extreme poverty among the peasants of Tuscany, as well as those living in the area of Venice. A diet of polenta, or corn mush – and very little else – was the usual fare of those who earned their living under the Italian system of *mezzadria*, or sharecropping. Their poverty meant little access to animal foods, and by the turn of the nineteenth century, pellagra had reached epidemic proportions in parts of Italy, especially in Lombardy, where it was estimated that from 5 to 20 percent of the population were its victims.

It was in northern Italy, where the peasants called the disease “the Rose,” that another new but lasting name, “pellagra” (*pelle plus agra*, or “rough skin”), was given to the illness by Italian physicians and other educated observers, who also developed new theories to explain its etiology and epidemiology. Some suggested that pellagra was caused by “insolation” or excessive exposure to the sun. Most, however, suspected that pellagra was caused by eating moldy

maize meal, and indeed, this belief persisted until the time of World War I.

Maize was a grain of the New World, unknown in Europe before the voyages of Christopher Columbus. At first, its food value was of little interest to Europeans, although one seventeenth-century herbalist, Caspar Bauhinus, cautioned against consuming too great a quantity of the new grain, as it produced an “itch.” But maize proved superior to other food crops, not only because of its higher yield of grain per acre but also because – as a new crop – it was not immediately subject to taxes, tithes, and feudal dues. By the end of the eighteenth century, it was grown in much of southern and eastern Europe, especially in Italy, Romania, and Spain.

It seems likely that by the time of Casal, pellagra was already endemic in parts of Europe. Political and economic trends had made peasants in both Italy and Spain even poorer, which forced an extremely limited diet on them. Interestingly, however, French peasants at that time, although equally poor, neither raised nor consumed as much maize as their southern neighbors. Consequently, it was only later, during the first third of the nineteenth century, that pellagra struck in France – at just about the time that that country’s expanding maize crop was becoming a more significant part of the diet.

Casal recognized that pellagra had some relation to diet: He was aware of its close association with maize consumption and knew that it could be cured by improving and varying what people ate. François Thiéry, a French physician who studied Casal’s still unpublished manuscript, announced these conclusions when he wrote the first published account of the disease, which appeared in 1755. Following the 1762 publication of Casal’s work, little was heard of pellagra in Spain until the 1840s. In that decade, however, a second French physician, Théophile Roussel, conducted an investigation and concluded that a number of differently named illnesses then plaguing Spain were all, in fact, pellagra. Soon afterward, Roussel began to work for the adoption of effective pellagra-preventing policies in his own country.

Africa

Throughout the rest of the nineteenth century, it became clear that pellagra was rampant in other parts of the world as well. During the 1890s, British epidemiologist Fleming Sandwith, working in Cairo, realized that endemic pellagra was all around him and began a systematic study of the disease. Like his predecessors, he found that pellagra was associated with a maize-based diet and extreme poverty (he added that residence in rural areas and exposure to the sun also seemed to be factors) and voiced agreement with the view that consumption of spoiled maize was the primary cause.

With his findings on endemic pellagra in Egypt, and a similar study he conducted of pellagra among

the Bantu people of South Africa during the Boer War, Sandwith had shown that the illness was by no means confined to Europe; in fact, at the end of the nineteenth century, it seemed obvious that pellagra had established far greater footholds elsewhere. His discovery was a significant accomplishment, but perhaps an even greater one was that through his reports and publications on pellagra (the first important ones to be written in English), he helped create awareness of the condition among physicians in the United States, who were unwittingly facing a pellagra epidemic of their own and would soon take the lead in fighting the disease.

The Americas

American Indians. It is generally accepted that at the time of Columbus's first voyage to the New World, maize cultivation was already widespread among the various American Indian groups (especially those of Central America and Mexico) and exerted a profound influence on their cultures. So important was this crop that a number of traditions described the origins of maize in miracle stories and creation myths (Roe 1973). Many such legends suggested "that man was formed from maize"; all saw maize as the "basic sustenance for the body and the soul" (Roe 1973: 10–11).

But despite the importance of maize in pre-Columbian culture and diet, the peoples of Mesoamerica seem not to have been troubled by pellagra. Indeed, their reliance on maize, while they apparently remained immune to the disease, posed a problem for early investigators advocating a link between maize-based diets and pellagra incidence (Roe 1973). But unlike European pellagrins, who consumed cornmeal mostly as mush, the Indians made their corn into tortillas – a process that involved soaking the maize grains in a lime solution, or perhaps liming them through treatment with campfire ashes. According to sixteenth- and seventeenth-century Spanish reports, such techniques were already in use at the time of European contact, and, in fact, archaeological evidence suggests that the process, virtually unchanged today, was first developed many centuries earlier (Roe 1973).

Many modern specialists in nutrition have suggested that the lime treatment of maize grain released its chemically bound niacin and made the vitamin available to the Indians, and some evidence indicates "that the amount of niacin so freed can be just sufficient to prevent pellagra" (Roe 1973: 15). But this explanation has not seemed sufficient to others, who have pointed to additional factors that operated (and still operate) to prevent the disease. Chief among these are the other niacin-containing vegetable foods that were eaten as well as maize, including squashes, chillies, and especially beans (both grown with maize and eaten with maize products). In addition, some sources of animal protein (usually rich in niacin), such as corn grubs and locusts, were consumed more or less regularly (Roe 1973).

Slaves, sharecroppers, and millworkers. Although the American Indians had historically not suffered from pellagra, other groups were not so fortunate. Even though physicians in the United States lacked knowledge of the disease, medical records permit the identification of a number of cases as far back as the early nineteenth century; in fact, it now seems clear that by the end of that century, pellagra was virtually epidemic, especially in the American South, and had been for some time. Southern agriculture, too long geared toward the production of cotton, had failed to diversify, with the result that the diet of the poorest classes of society lacked much in the way of variety (Etheridge 1972, 1993; Roe 1973; Beardsley 1987). Pellagra seems to have been a periodic visitor to slave cabins during antebellum decades and, later, plagued the southern poor generally; however, it was frequently ignored or – because of its protean symptoms – confused with other diseases (Kiple and Kiple 1977).

In the early years of the twentieth century, however, a pellagra epidemic occurred among mental patients at the Alabama Institution for Negroes, and unlike previous incidents, this outbreak received a measure of publicity. Moreover, it appears that this was not the first time physicians at the institution had seen the disease, and two of them, George H. Searcy and Emit L. McCafferty, produced a report on 88 pellagra patients, 57 of whom had died. Published in the *Journal of the American Medical Association* in 1907, this study provided the kind of information that helped other physicians to recognize the disease (Roe 1973). In the months that followed, thousands of cases were diagnosed, most of them in southern states.

Miracle "cures" for pellagra began to appear, and physicians, despite the debatable effectiveness of such "medicines" and the danger posed to their patients, made liberal use of "Fowler's solution," "Atoxyl," "Salvarsan," and other arsenical preparations. They also employed purges, antiseptic injections, various tonics, blood transfusions from recovered pellagrins, and, in one case, "treatment" with static electricity. A U.S. congressman referred to the latter "cure" as "simply marvelous" (Roe 1973: 95).

Also much in evidence were patent medicines such as "Ez-X-Ba River, The Stream of Life," which was sold for 5 dollars a bottle by the Dedmond Remedy Company, founded in 1911 by Ezxbda Dedmond. This product was reputed to cure pellagra within weeks, and early in its history the company claimed hundreds of cures. Another such was "Pellagracyde," made by the National Pellagra Remedy Company. Both of these firms, based in South Carolina, were challenged in 1912 by that state's Board of Health, and their products were attacked as fraudulent by the American Medical Association. Analysis of the concoctions revealed no active elements at all, but the companies remained in business. In 1913, "Doctor" Baughn's

American Compounding Company followed much the same course with its "Baughn's Pellagra Remedy" in Alabama. The quack medicines enjoyed much more popularity and confidence among the general public than did the treatments offered by legitimate physicians; indeed, the first encounter many poor pellagra victims had with a medical doctor was when they were near death and committed to a mental hospital (Roe 1973).

Except for institutionalized people (frequently victims of nutritional deficiency diseases because of a restricted diet), southerners suffering from pellagra were generally to be found among the rural poor, many of whom eked out a living by sharecropping, or among the workers in textile mills and their families. It could be said that the millworkers themselves "were in a way institutionalized" in areas where much of the housing, employment, food sources, and medical care were controlled by the firms operating the mills. Indeed, the millworkers' situation was perhaps fractionally worse than that of their rural counterparts, who sometimes had greater access to what little fresh food was produced (Beardsley 1987: 56-7).

Following the epidemic at the mental home in Alabama, the U.S. Public Health Service entered the battle against pellagra, and beginning in 1914, epidemiologist Joseph Goldberger succeeded in curing institutionalized pellagrins by altering their diets. His next trial, also successful, was to induce the disease in prison inmates who volunteered to subsist on a restricted diet. In addition, he proved through experiments on himself and his colleagues that pellagra was neither transmittable nor infective.

Moreover, Goldberger initiated a multiyear study of the disease which revealed that poverty and lack of dietary variety were consistently associated with pellagra in the American South. This study, which later attained renown as a classic model of epidemiology, showed conclusively that pellagrous millworkers and agricultural laborers were on the lowest rung of the economic ladder, and that the traditional diet of the southern poor - cornmeal bread, "fatback" pork, and molasses - was all the food that most such people had access to, or could afford. Thus, poverty, manifesting itself through bad nutrition, had resulted in pellagra attaining epidemic proportions among the lowest-paid classes in the society.

Goldberger had proved that the cause of pellagra was a dietary deficiency and, furthermore, had demonstrated that the deficiency - and therefore the disease - resulted from the prevailing social and economic conditions. Further dietary tests indicated that the diet of pellagrins lacked some crucial element, which was referred to as the pellagra-preventing factor (the "P-P factor" for short), and that this substance, though yet unidentified, was present in meat and dairy foods and some vegetables.

A Disease of an Inferior Social Group

Class inequalities have been used from early times to account for variances in the prevalence of endemic diseases among social groups. Yet the reasons advanced to explain why the poor are more susceptible to certain diseases have changed over the years. In the past, their disease vulnerability was blamed on "bad blood," unclean habits, angry gods, and - somewhat more correctly - the fact that they lived in close proximity to their animals. Today, the blame is placed on an unwillingness to seek preventive health care or to eat the type of diet known to reduce disease risks.

In the case of pellagra, almost all major contributors to the literature of the past subscribed to the view that it was found only among indigent people. Indeed, "the Rose" was considered to be a "brand" of extreme poverty. However, as with other diseases, explanations of why the poor contracted pellagra - and more affluent people did not - have varied.

Hereditary Weakness and Susceptibility

From the time that pellagra was first described until the end of World War I, the occurrence of multiple cases of pellagra in a single family was explained as a hereditary weakness within that family. That pellagra was a manifestation of "bad blood" was a common extrapolation of the beliefs of the Eugenists. At the height of the Eugenics movement, it was believed that children with "bad" traits were born of marital unions between parents who carried these predispositions. Such a theory might be considered a forerunner of human genetics. However, Eugenists believed that "bad blood" was a function of moral traits as well as of traits relating to physiognomy and disease susceptibility. Clearly underpinning the Eugenists' apparently scientific explanation of the pellagra trait was the traditional association of moral weakness and disease. The origin of the concept of "bad blood" lies in the idea that evil is inherited (Ricoeur 1967).

G. K. Chesterton wrote in his book *Eugenics and Other Evils* (1922) that H. G. Wells should be given a medal as the Eugenist who destroyed Eugenics. Wells's argument (1903) was that we cannot be certain of the inheritance of health because health is not a quality but rather a comparative term. Yet despite Chesterton's assertion that Wells put the Eugenists' theories to rest, the idea of "bad blood" persisted and has been frequently used to refer to the taint of syphilis as well as that of pellagra. Moreover, disease vulnerability was somehow supposed to be related to inferior social position. Curiously, physicians who viewed pellagra as the result of inferior parentage could also think in terms of traits that were risk factors for the disease - these same traits also being risk factors for criminal behavior and alcohol abuse.

Diatheasis and Biotype

Jean-Marie Gustave Hameau (1853) wrote that “[l]a pellagre est une diathèse particulière,” a statement that reflects the viewpoint of many observers of the disease from the late eighteenth through the nineteenth century. The people considered susceptible to pellagra were indigent peasants whose vulnerability was somehow linked to their lifestyle.

A corollary idea was that specific “biotypes” were susceptible to pellagra. Dr. Charles Davenport, for example, believed that pellagra was a communicable disease, but only among people of particular biotypes. In a paper published in 1916, he wrote of the risk of acquiring pellagra: “It appears that certain races or blood lines react in the pellagra families in a special and differential fashion.” He also believed that the type of constitution a person possessed determined the progress of the disease (Davenport 1916).

Criminality and Hereditary Pellagra

Pellagra was also thought to occur more frequently among the criminal classes, or at least among those with the potential for wrongdoing. Thus, particularly in the nineteenth century, the social environment of pellagrins was believed to play a definitive role in explaining their disease.

Cesare Lombroso, who became professor of psychiatry at Pavia in 1862, centered his research on relationships between mental and physical disorders. He thought that criminals could be identified by certain physical characteristics. He was also convinced that there were intrinsic physical and moral characteristics that explained an individual’s susceptibility to toxins in moldy maize grain, which (he believed) caused pellagra (Lombroso 1869, 1893).

Lombroso thought in terms of a hereditary pellagra that existed in both a mild and a severe form. Hereditary pellagra, he wrote, could be recognized in the second year of life when it was manifested by pain, indigestion, a voracious appetite, diarrhea, and cachexia. Moreover, those afflicted with hereditary pellagra also suffered physical anomalies such as faulty development of the skull with brachycephaly or dolichocephaly, a receding forehead, badly set ears, asymmetry of the face, and abnormalities of the external genitalia (Lombroso 1893).

Pellagra and Alcohol Abuse

Once pellagra had been noted as endemic in the United States, it was observed to be common among those who drank to excess, and the term “pseudo-pellagra” was sometimes used to describe the disease when it occurred in alcoholics. Such a term implied only that these patients had a pellagra-like disease, but eventually it was recognized that genuine pellagra was particularly common among alcoholics, although the reasons for their vulnerability were not understood.

In 1928, J. V. Klauder and N. W. Winkleman reported

their studies of pellagra among alcoholics and stressed that the disease was most likely to occur in chronically heavy drinkers who had been on binges lasting several weeks. The researchers also made the observation that during these “debauches,” the alcoholics ate very little food, and usually only food of one type – soup. The soup in question was, no doubt, frequently a handout of soup kitchens and not likely to be laden with nutrients. Thus, in the case of some alcoholics, pellagra vulnerability was to some extent linked to a dependence of the have-not group on donated food provided as charity.

N. Jolliffe (1940) was the first nutritionist to observe that there were different reasons why alcoholics so often developed pellagra. He suggested that there were four causal relationships between the disease and alcohol abuse. First, the gastritis associated with heavy alcohol consumption could lead to poor food intake. Second, alcohol-related changes in the gastrointestinal tract could interfere with vitamin absorption. Third, alcohol might be substituted for foods containing vitamins. And last, there was probably an increased vitamin requirement in alcoholics, which resulted from alcohol ingestion.

But despite the possibility of a scientific explanation for the prevalence of pellagra among alcohol abusers, both physicians and the lay public were – and often still are – of the opinion that pellagra strikes alcoholics because of their bad health habits in general (and, perhaps, because of their “immoral ways” in particular). In a 1911 review of cases of pellagra, Dr. Beverley Tucker of Richmond, Virginia, discussed the “pernicious habits” of 15 of her 55 patients with the disease. The habits she considered “pernicious” included the abuse of alcohol, tobacco, and opium.

Revelation of the Social Causes of Pellagra

In the United States, support for the concept that endemic pellagra was a result of extreme poverty came from the studies carried out by Goldberger and his colleagues, G. A. Wheeler and the economist Edgar Sydenstricker. These individuals comprised the first American team to employ epidemiological methods and economic analysis to explain the prevalence and distribution of pellagra in the South. In their 1916 study of cotton-mill villages in South Carolina, the investigators found that the proportion of families with pellagra declined as income increased. Those families with more than one case were always within the lowest income group.

Poor hygiene and sanitation, as well as differences in age and gender distribution – all of which had been considered significant factors – were only gradually eliminated as causes. It became clear that, in general, households with higher incomes and access to food sources outside the factory commissaries, including the produce of home-owned cows, pigs, and poultry,

enjoyed a much lower incidence of pellagra. Thus, buying power and access to pellagra-protective foods conferred disease “immunity” (Goldberger, Wheeler, and Sydenstricker 1918, 1974). Sydenstricker also showed in a later study that family income had to be very low before the standard of living forced consumption of a pellagra-producing diet (Sydenstricker 1933; Roe 1973).

Eradication Policies in France and the United States

Soon after its discovery (by the mid-eighteenth century), pellagra was also explained as either an outcome of exposure to extreme climates or a product of contagion. In the mid-nineteenth century, Daniel Drake (1850) grouped causes of disease into three classifications. These included telluric or geological, climatic or meteorological, and social or physiological influences. Later in the century, diseases were also found to be caused by agents such as light, food toxins, and pathogens. As each of these agents of disease was recognized, it became a suggested cause of pellagra. Infection, for example, was popular as a causative agent because it seemed plausible that poor sharecroppers, living in fly-infested dwellings, might acquire the supposed “germ” from their surroundings.

Yet the applicability of the germ theory to pellagra could not be demonstrated, and – especially as nutrient deficiencies had been implicated in other diseases, like scurvy – the interest of researchers returned to the diet of the afflicted. But partly because of deliberate attempts to downplay the poverty of workers, and partly because of the reluctance of medical practitioners to agree on a nutritional explanation, it took a long time for the etiology of pellagra to be understood.

In an interesting observation, S. J. Kunitz (1988), in a paper dealing with hookworm infestation and pellagra, remarked that past explanations of these diseases were influenced by investigators’ ideologies and values rather than being derived from analyses of the intrinsic nature of the illnesses. Kunitz further suggested that such biases also conditioned approaches to disease prevention and cure.

Perhaps such influences can be seen in the totally different public-health approaches to pellagra eradication adopted in France in the 1840s and in the United States in the 1940s. In France, Roussel, without any knowledge of the biological cause of pellagra but with a keen observer’s eye for the social environment, urged the French government to drain the salt marshes in southwestern France, where the disease was most prevalent, so that a diversity of food crops could be grown and animals raised. More generally, he was also remarkably successful in getting the French government to change the country’s agricultural system in ways that contributed to the health of the inhabitants of the rural regions (Roussel 1845, 1866).

In the United States, Conrad A. Elvehjem of the University of Wisconsin discovered the pellagra-preventing factor, first called nicotinic acid and later named niacin. This “breakthrough” occurred in 1937 as a result of his study of blacktongue, the canine equivalent of pellagra (Elvehjem et al. 1938), and Elvehjem was subsequently influential in promoting the enrichment of bread and cereal grains with niacin in an effort to bring the career of pellagra to a close. The mandatory enrichment program was instituted in 1943, following the recommendation of the Food and Nutrition Board, of which Elvehjem was a member (Wilder and Williams 1944).

Yet, as T. H. Jukes (1989) has pointed out, however successful the bread- and cereal-enrichment program was in preventing endemic pellagra, it was already generally accepted by the 1940s that pellagra could be prevented – and also cured – by what he termed a “good” diet; in this context, a “good” diet was one that included milk and meat. The government, however, chose not to concentrate on changing the living conditions of the disadvantaged, such as sharecroppers and millworkers, but rather adopted a policy of providing them with cheap, niacin-containing staple foods.

Thus, neither the sharecropping system nor pellagra were ended by enlightened public policy. Although the beginning of the end of pellagra in the United States lay in the expedient of food fortification, its ultimate demise can only be found in the economic events of the 1930s and 1940s, which, by bringing greater affluence to the South, both eliminated pellagra as a major health concern and spurred the end of sharecropping (Roe 1974).

Endemic Pellagra in Refugee Camps

Pellagra, however, is far from dead in the developing world and is often seen in the midst of chaotic situations. For example, the disease surfaced a few years ago in Malawi, when thousands of Mozambicans fled the civil conflict in their own country to seek refuge there. Once in Malawi, they lived in refugee camps or nearby villages, where – between July and October 1989 – 1,169 cases of pellagra were diagnosed among refugees living in 11 sites. From February 1 through October 30, 1990, another 17,878 cases were reported among a population of 285,942 refugees; in other words, over 6 percent were afflicted. But the rate of affliction varied from one location to another, ranging from 0.5 percent to 13.2 percent. Moreover, females were more than seven times as likely to be afflicted as males; young children, however, were substantially less affected than adults, as has generally been the case during outbreaks of pellagra. The disease was also less common among those who lived in integrated villages rather than in camps (Editorial 1991).

French epidemiologists working for *Médecins*

Sans Frontières, who investigated the epidemic, found that those refugees who escaped pellagra were more likely to have gardens, have a daily supply of peanuts (an important niacin-containing staple of the region), or have the ability to mill maize. At the time of the epidemic, peanut distribution had been disrupted, and the maize sent in by donor nations was neither vitamin-enriched nor even ground into meal. Thus, those who developed pellagra were totally dependent on the maize ration, and the greater vulnerability of women was explained by the tendency of males to appropriate nuts, meats, and fish (foods high in niacin as well as tryptophan, its precursor) for themselves.

Clearly, the appearance of a major epidemic of pellagra toward the end of the twentieth century – when the means of preventing the disease have been known for more than 50 years – suggests a substantial error in judgment by supposedly compassionate nations.

Lessons to Be Learned

At the risk of belaboring points that have already been made, by way of conclusion it seems worthwhile briefly to revisit some past notions about pellagra. The first of these is that pellagra was the fault of the afflicted, rather than of those who maintained the existing inequalities of the social system. The second (not all that different from the first) was that the eating habits of pellagrins were the consequence of an unwillingness to change their lifestyles.

Some social critics of the past did better than the scientists in understanding pellagra. French novelist Edmond About (1858), for example, clearly indicated his grasp of the social causes of the disease when he had one of his characters remark that pellagra would continue to exist in the marshy southwestern area of his country (the Landes region) until the nature of the environment changed. His succinct prediction,

*Tant que Lande sera lande,
La pellagre te demande*

(As long as the Landes remains a moor,
There pellagra will claim the poor),

suggests that pellagra was the fault of the society in which it raged rather than that of the peasant who suffered the deprivations of such a society.

In rural areas of the southern United States, the sharecroppers' "bad habit" of eating an unvaried diet of cornmeal with occasional fatback and molasses – the "three Ms": maize, molasses, and meat (but only the fatback type of meat) – was understood to increase their nutritional risk of pellagra. In urban environments, the "bad" food habits of alcoholics were thought to explain their pellagra susceptibility. Lost in this tendency to blame the victim was the fact that sharecroppers ate a deficient diet because they had little or no access to food other than the "three

Ms." Similarly, pellagra susceptibility in "down-and-out" alcoholics can be explained by society's inability to accept alcoholism as a disease and the consequent belief that alcoholics should not receive adequate nutritional assistance because this would only encourage them to continue in their lifestyle.

In the case of refugee camps, not only were the refugees viewed as an inferior social group, but (as with the alcoholics who were the urban pellagrins of the 1920s and 1930s) they were fed with indifference by donor nations without any effort to improve their health and quality of life. Even today, nations that send food to those living in such camps assume little responsibility for providing an adequate diet for the recipients. Rather, they continue to ship food – like maize – that is unfamiliar to the consumers and is grossly deficient in nutrients.

Certainly, past – often elitist – views of pellagra as the fault of the pellagrin are no longer acceptable today. Moreover, instead of claiming the conquest of endemic pellagra as a scientific triumph, we might ask why pellagra came about in the first place and why it persisted for so long. Finally, we should ask why, even after the means of pellagra prevention are fully understood, there are still serious outbreaks of the disease in various parts of the world.

*Daphne A. Roe
Stephen V. Beck*

Thanks to Dr. Charles C. Dickinson, III, for his help with the quotation from the work of Edmond About.

Bibliography

- About, E. 1858. *Maitre Pierre*. Second edition. Paris.
- Albera, G. M. 1784. *Trattato teorico pratico delle malattie dell' insolato di primavera volgarmente dette della pellagra*. Varese, Italy.
- Bateman, T. 1813. *A practical synopsis of cutaneous diseases*. London.
- Beardsley, Edward H. 1987. *A history of neglect: Health care for blacks and mill workers in the twentieth-century South*. Knoxville, Tenn.
- Black Report. 1982, 1988. In *Inequalities of health*, 33–42. London.
- Carpenter, K. J., and W. J. Lewin. 1985. A reexamination of the composition of diets associated with pellagra. *Journal of Nutrition* 115: 543–52.
- Casal, G. 1762. De affectione quae vulgo in hac regione "mal de la rosa" nuncupatur. In *Historia natural y medica de el Principado de Asturias: Obra posthuma*, Vol. 3, 327–60. Madrid.
- Chesterton, G. K. 1922. *Eugenics and other evils*. London.
- Davenport, C. B. 1916. The hereditary factor in pellagra. *Archives of Internal Medicine* 18: 32–75.
- Drake, D. 1850. *A systematic treatise, historical, etiological and practical on the principal diseases of the interior valley of North America*. Cincinnati, Ohio.
- Editorial. 1991. Outbreak of pellagra among Mozambican refugees – Malawi, 1990. *Monthly Morbidity and Mortality Report* 40: 209–13.

- Elvehjem, C. A., R. J. Madden, F. M. Strong, and D. W. Woolley. 1938. The isolation and identification of the anti-black-tongue factor. *Journal of Biological Chemistry* 123: 137–49.
- Etheridge, Elizabeth W. 1972. *The butterfly caste: A social history of pellagra in the South*. Westport, Conn.
1993. Pellagra. In *The Cambridge world history of human disease*, ed. Kenneth F. Kiple, 918–24. Cambridge and New York.
- Frapolli, F. 1771. *Animadversiones in morbum vulgo pellagram*. Milan.
- Frye, N. 1981. *The great code*. New York.
- Goldberger, J., G. A. Wheeler, and E. Sydenstricker. 1918. A study of the diet of non-pellagrous and pellagrous households. *Journal of the American Medical Association* 71: 944–9.
1920. A study of the relation of diet to pellagra incidence in seven textile-mill communities of South Carolina in 1916. *Public Health Report* 35: 648–713.
1974. A study of the relation of family income and other economic factors to pellagra incidence in seven cotton mill villages in South Carolina in 1916. In *The challenge of facts: Selected public health papers of Edgar Sydenstricker*, ed. R. V. Kasius, 263–313. New York.
- Goldsmith, G. A. 1956. Studies on niacin requirements in man. II: Comparative effects of diets containing lime-treated and untreated corn in the production of experimental pellagra. *American Journal of Clinical Nutrition* 4: 151–60.
- Goldsmith, G. A., H. P. Sarett, U. D. Register, and J. Gibbens. 1952. Studies of niacin requirement of man. I. Experimental pellagra in subjects on diets low in niacin and tryptophan. *Journal of Clinical Investigation* 31: 533–42.
- Guthrie, D. 1945. *A history of medicine*. London.
- Hameau, J.-M. G. 1853. *De la pellagre*. Paris.
- Hauerwas, S. 1975. *Character and the Christian life*. San Antonio, Tex.
- Jolliffe, N. 1940. The influence of alcohol on the adequacy of the B vitamins in the American diet. *Quarterly Journal of Studies on Alcohol* 1: 74–84.
- Jukes, T. H. 1989. The prevention and conquest of scurvy, beriberi and pellagra. *Preventative Medicine* 18: 877–83.
- Kiple, Kenneth F., and Virginia Himmelsteib King. 1981. *Another dimension to the black diaspora: Diet, disease, and racism*. Cambridge.
- Kiple, Kenneth F., and Virginia H. Kiple. 1977. Black tongue and black men: Pellagra and slavery in the antebellum South. *The Journal of Southern History* 43: 411–28.
- Klauder, J. V., and N. W. Winkleman. 1928. Pellagra among chronic alcoholic addicts. *Journal of the American Medical Association* 90: 364–71.
- Kunitz, S. J. 1988. Hookworm and pellagra: Exemplary diseases in the New South. *Journal of Health and Social Behavior* 29: 139–48.
- Lombroso, C. 1869. *Studii clinici ed sperimentali sulla natura causa e terapia della pellagra*. Bologna.
1893. Sull'etiologia e sulla cura della pellagra (1892). *Lavori di Congresso di Medicina*, 160–86. Milan.
- Marie, A. 1908. *La pellagre*. Paris.
- Parsons, T. 1972. Definitions of health and illness in the light of American values and social structure. In *Patients, physicians and illness*, ed. E. G. Jaco, 97–117. New York.
- Ricoeur, P. 1967. *The symbolism of evil*. Trans. Emerson Buchanan. Boston, Mass.
- Roe, D. A. 1973. *A plague of corn: The social history of pellagra*. Ithaca, N.Y.
1974. The sharecroppers' plague. *Natural History* 83: 52–63.
1991. Pellagra. In *Hunter's tropical medicine*. Seventh edition, ed. G. T. Strickland, 932–4. Philadelphia, Pa.
- Roussel, T. 1845. *De la pellagre*. Paris.
1866. *Traité de la pellagre et de la pseudopellagre*. Paris.
- Samuel, Viscount. 1949. *Creative man and other addresses*. London.
- Strambio, G. 1785–9. *De pellagra observationes in regio pella-grosorum nosocomio factae a calendis junii anni 1784, usque ad finem anni 1785*. 3 vols. Milan.
- Sydenstricker, E. 1933. *Health and environment*. New York.
- Sydenstricker, V. P., H. L. Schmidt, M. C. Fulton, et al. 1938. Treatment of pellagra with nicotinic acid and observations in fifty-five cases. *Southern Medical Journal* 31: 1155–63.
- Tucker, B. R. 1911. Pellagra, with the analytical study of 55 non-institutional or sporadic cases. *Journal of the American Medical Association* 56: 246–51.
- Underhill, F. P. 1932. Clinical aspects of vitamin G deficiency. *Journal of the American Medical Association* 99: 120–4.
- Wells, H. G. 1903. *Mankind in the making*. Leipzig.
- Wilder, R. M., and R. R. Williams. 1944. *Enrichment of flour and bread: A history of the movement*. Washington, D.C.

IV.D.6 🐉 Pica

Broadly stated, pica is the term given to the compulsive consumption of substances not generally considered food. However, a precise definition of pica is somewhat elusive because understandings of what constitutes “food,” what symptoms signify pica, and explanations of what causes the condition vary with historical and cultural context. A more specific definition of pica is “the chronic, compulsive eating of non-foods such as earth, ashes, chalk, and lead-paint chips . . .” (Hunter 1973: 171), but it may also include a “false or craving appetite” or “deliberate ingestion of a bizarre selection of food,” as well as the compulsive ingestion of nonnutritive or nonfood items such as ice and ice water (Parry-Jones and Parry-Jones 1994: 290).

Pica, in various forms, has been widely noted historically and geographically, primarily in medical texts and anthropological writings (see, for example, Laufer 1930; Cooper 1957; Anell and Lagercrantz 1958). Its practice, although not considered a disease, is of medical concern because ingestion of some substances may result in disease. Additionally, there are types of pica that have been linked by medical researchers to the correction of mineral deficiencies (see, for example, Coltman 1969; Crosby 1971; Hunter 1973). Pica is classified by the DSM-III-R (American Psychiatric Association 1987) and the ICD-10 of the World Health Organization (1992) as an eating disorder, along with anorexia nervosa, bulimia, and infant rumination. Various forms of pica have also been associated with mental retardation.

The incidence of pica in any particular population is difficult to determine because of a tendency to conceal eating behavior that may be considered abnormal or deviant within the cultural context. In addition, the varying definitions of pica we have taken note of also contribute to difficulty in documenting the scope of the behavior. Thus, although pica has been widely observed (both historically and geographically) and has been the subject of much research in many disciplines, it remains poorly understood.

Classifications

Pica is generally classified according to the type of substance consumed. Names for subclassifications of pica are comprised of the Greek word for the ingested substance and the suffix from the Greek word “phagein,” meaning “to eat” (Moore and Sears 1994). Cross-culturally, the most commonly noted and explored type of pica is geophagy or geophagia, the consumption of earth and especially clay. Other types include ingestion of ice or ice water (pagophagia); laundry starch (amylophagia); hair (trichophagia); gravel, stones, or pebbles (lithophagia); leaves, grass, or other plants (foliophagia); feces (coprophagia); and unusual amounts of lettuce (lectophagia), peanuts (gooberphagia), and raw potatoes (geomelophagia). Paint, plaster, coal, chalk, cloth, pepper, coffee grounds, paper, cigarette butts, and other household items are also commonly consumed by those engaged in pica (Feldman 1986: 521).

There is some agreement historically and cross-culturally that the populations most prone to pica are young children, pregnant women, persons with mental illness, and the mentally retarded. There is also similarity cross-culturally in the types of items most frequently consumed. These include coal, ice, chalk, plaster, and various types of earth, in particular clay.

As a rule, children who engage in pica are under the age of 6 (Castiglia 1993). The things they consume are restricted by proximity to their grasp and normally tend to be relatively harmless items such as cloth, dirt, leaves, sand, and small rocks or pebbles (Parry-Jones and Parry-Jones 1992). Children who chew furniture or eat paint or plaster can be harmed, and plumbophagia, the ingestion of lead paint, is an important cause of lead poisoning. In fact, the practice of plumbophagia has led to the banning of lead-based paints for interior use in homes.

Trichophagia refers to the ingestion of hair and is one of the types of pica found most often among children. It is especially associated with the habit of girls chewing on long hair and is believed to be related to other, somewhat common, behaviors such as chewing one's fingernails and chewing pencils (Higgins 1993). Trichophagia is of medical concern when substantial amounts of hair result in the formation of “hair balls” within the intestinal tract.

History

The term “pica” was first coined by the French physician Ambroise Paré in the sixteenth century, although references to the syndrome predated him by centuries. Aristotle and Socrates both wrote about “earth eating,” and during the classical periods of Greece and Rome, red clay lozenges from Lemnos were believed by physicians such as Galen to be antidotes for poison and cures for illnesses. They were also believed to facilitate childbirth. The lozenges were called *terra sigillata* (sealed earth) and stamped with the seals of the goddesses Artemis and Diana. As Christianity spread, these seals were replaced with Christian symbols and the lozenges, blessed by monks, were traded throughout western Europe and the Mediterranean region with the approval of the Roman Catholic Church (Hunter and De Kleine 1984).

The word “pica,” and its older variant, *cissa*, come from the Latin word for “magpie,” a bird thought to have a not very discriminatory appetite where edible and nonedible substances were concerned. Nineteenth-century medical texts describe both the eating behavior of magpies and that of humans with pica as consisting of an appetite for unusual edible and nonedible items (see, for example, Hooper 1811). The misconception that the magpie consumed earth and clay was likely based on observations of magpies collecting clay to build nests.

Pica was classified in Greek and Roman medical texts as a form of morbid or depraved appetite. In 1638, M. H. Boezo distinguished pica, the consumption of nonfoods, from “malacia,” a voracious appetite for “normal” foods. He attributed malacia in pregnant women to mental changes thought to occur in pregnancy. (As early as the sixth century A.D., pica was thought to result from the cessation of menstruation during pregnancy [Cooper 1957]). Today, malacia, or the craving and binge eating of specific foods, is considered a form of pica (Castiglia 1993). (For the early literature on pica, see Cooper 1957 and Halsted 1968.)

In mid-sixteenth-century England, pica was associated with coal eating among pregnant women and children. But within western Europe and the United States from the sixteenth century through the late nineteenth century, pica was commonly understood as the consumption by young women of substances such as lime, coal, vinegar, and chalk so as to achieve a pale complexion and otherwise improve on appearance (Parry-Jones and Parry-Jones 1994: 290). Historically, this condition was said to be accompanied by “chlorosis” or “green sickness” in prepubescent girls and young women. Chlorosis, a disease recognized from the sixteenth century through the late nineteenth century, was characterized by a loss of menses or irregular menstruation and was accompanied by symptoms such as listlessness, pallid skin, loss of appetite, and weight loss.

It is interesting to note that the debate in early medical literature about the causes of chlorosis prefigures the current debate about pica and iron deficiency in terms of cause and effect. For example, the consumption of nonfoods by young women in order to achieve a pale complexion could easily have resulted in iron-deficiency anemia or chlorosis. However, iron-deficiency anemia can cause cravings for nonfoods, and chlorotic females ingested large amounts of unusual foods such as pepper, nutmeg, and raw corn, as well as nonfoods such as plaster. In addition, psychological reasons such as sexual frustration and nervous conditions were considered possible causes of both chlorosis and pica (Loudon 1980). In the twentieth century, pica among young women has been manifested by excessive consumption of real foods such as fruit and vegetables and nonfood substances like ice (Parry-Jones and Parry-Jones 1992). (For an extensive historical account of pica within Western industrialized cultural contexts, see Parry-Jones and Parry-Jones 1994).

In various regions of the world, especially in tropical zones, pica most often takes the form of geophagy. Harry D. Eastwell has noted that geophagy is associated with the "world's poor or more tribally oriented people" (1979: 264). Other investigators have characterized such groups as constituting "subsistence" societies (Hunter 1973), although in this chapter, the term "nonindustrialized" societies is used. In these societies, geophagy has been observed for many centuries and variously attributed to religious, cultural, and physiological causes (Hunter 1973; Hunter and De Kleine 1984; Parry-Jones and Parry-Jones 1992). Geophagy, or "dirt eating," was also thought to be a peculiar affliction of enslaved Africans and, later, lower-income African-Americans and whites in the southern United States, who were characterized as "dirt eaters" (Forsyth and Benoit 1989).

Etiology

Explanations of pica are numerous, reflecting the diversity of items consumed and the geographical regions within which the compulsive consumption occurs. In addition to the historical and cultural practices just mentioned, explanations also include psychiatric disorders and psychological and physiological needs, satisfaction of oral needs, behavioral disorders, responses to physiological or psychological stress, and the use of nonfoods for medicinal or pharmacological purposes (see, for example, Talkington et al. 1970; Hunter 1973; Crosby 1976; Eastwell 1979; Hunter and De Kleine 1984; Prince 1989; Horst 1990; Reid 1992). In fact, the practice of pica is so widespread both geographically and historically that one might be tempted to question its characterization as an abnormal practice.

Nonetheless, at least in the West, pica is considered an aberrant behavior warranting medical or psycho-

logical treatment. Within the medical literature, pica is discussed in terms of possible biological causes and their negative consequences. The psychological literature characterizes pica as a pathological behavior linked to other eating disorders or found among populations of children and the mentally retarded.

However, this tendency to view pica as a pathological eating practice, an idiosyncrasy of "primitive" peoples, or an affliction primarily of the rural and impoverished has resulted in a failure to recognize the nutritional, medicinal, and cultural importance of geophagy in nonindustrialized cultural contexts.

Geophagy (Geophagia)

Geophagy, or the consumption of earth substances, is the most widely observed and researched type of pica. The term "geophagy" was first coined by Aristotle and means "dirt eating." Geophagia has also been termed *allotriophagia* by Sophocles, *erdessen* in medieval Germany, *mal d'estomac* in French, *citta* in Latin, and *cachexia africana* - literally meaning a "wasting away of Africans" - a phrase employed by slave owners and physicians in the West Indies and southern United States.

Despite the historical evidence for various forms of pica throughout the world, geophagy is the only type to be discussed extensively from a cross-cultural perspective, especially in ethnographic anthropological and geographical literature. In transcultural perspective, both anthropologists and geographers have described it as a socially acceptable custom, with specific cultural meanings and functions. Such literature reports geophagy to be most frequently practiced in tropical areas of Africa, Latin America, and the Caribbean, as well as in the southern United States. It is also practiced widely in parts of Iran, India, and China, and in tropical areas of Indonesia and Oceania (Anell and Lagercrantz 1958; Hunter 1973).

In the Andes Mountains of Peru and Bolivia, two dozen comestible or edible earth substances have been found listed in pre-Columbian Incan sources, and about six different comestible earths have been discovered archaeologically in pre-Incan contexts. In fact, evidence that the practice of eating earth is at least some 2,500 years old in the Andean region was discovered when a specimen of comestible earth was recovered from a Bolivian site dating from 400 B.C. (Browman and Gundersen 1993). Although cultural explanations have been popular, physiological explanations of geophagy as an adaptive human behavior are also important (Hunter 1973; Hunter and De Kleine 1984; Johns and Duquette 1991; Browman and Gundersen 1993).

Some of the first domesticated plants contained substances toxic to humans and were treated in processes involving earth substances to make them less toxic. The absorptive properties of clay were well known to ancient physicians who included *terra sig-*

illata and *terra silesiaca* in their pharmacopoeias for the treatment of poison (Dannenfeldt 1984). Several authors note the example of a condemned man, during the sixteenth century, who elected to swallow a lethal dose of mercury if he were first allowed to ingest clay. He reportedly survived an amount of mercury three times the normal lethal dose and was granted a pardon for having contributed to the medical knowledge of the time (Halsted 1968; Dannenfeldt 1984). In our times, kaolinite is the common type of clay used in medicines and the primary ingredient in commercially marketed Kaopectate.

Clay is also used medicinally in nonindustrialized societies where hookworm is a common ailment. This intestinal parasite causes gastric distress that is frequently alleviated with clay. Clay is also employed as a treatment for diarrhea, heartburn, and intestinal gas and has been used to relieve nausea and vomiting in pregnant women (Anell and Lagercrantz 1958; Hunter 1973). But authorities caution that the practice of geophagy can also introduce intestinal parasites into the body (Castiglia 1993) as well as cause intestinal blockage and excessive wear of dental enamel.

John M. Hunter (1973) hypothesized that a connection exists between mineral deficiencies (particularly those resulting from increased nutritional requirements during pregnancy) and the cultural practice of geophagy in Africa. Moreover, along with Renate De Kleine, he suggested that clay eating in Central America may be a “behavioral response to a physiological need” created by various mineral deficiencies, particularly during pregnancy (Hunter and De Kleine 1984: 157). Similarly, Donald E. Vermeer investigated geophagy among the Tiv of Nigeria and the Ewe of Ghana and discovered that clays were consumed during pregnancy as a treatment for diarrhea as well as for the minerals that they contain (Vermeer 1966; Vermeer and Frate 1975).

Hunter, who described geophagy in Africa as “common among children and adults” (1973: 171), acknowledged that the practice also has a cultural basis. For example, earth, taken from a shrine or holy burial site, is eaten for religious purposes or to swear oaths. However, earth eating is most commonly viewed as a remedy. Syphilis, diarrhea, and gastrointestinal discomfort caused by parasitic diseases (such as hookworm) are all conditions treated with geophagy by the general population, but in Africa, the practice of consuming clay is most frequently found among pregnant women. In a field study conducted in the Kailahun District of Sierra Leone, Hunter (1984) found that 50 percent of pregnant women ate clay from termite mounds and 7 percent ate vespid mud or clay from mud-daubing wasps’ nests. In both cases, the mud was cooked over a fire until dried and blackened. An analysis of the clay to determine mineral content and availability for humans led Hunter to conclude that the practice of geophagy was “sensible and appropriate behavior” – as the clays made a signif-

icant contribution to the calcium, manganese, iron, and other mineral requirements of pregnant women (1984: 11).

In other parts of Africa, well-known and highly regarded clays are “extracted, processed, and passed from producers through middlemen to retailers and reach a wide consumer public through a network of periodic markets” (Hunter 1973: 173). Analysis of the nutritional content of a sample of such clays from Ghana revealed that clays of distinctive shapes (indicating a sort of brand) provided mineral supplementation in distinctive, varying proportions. For example, Hunter wrote:

a distinctive end-to-end, truncated cone-shaped clay (sample six), is richest in two macronutrients, potassium and magnesium, but is poor in calcium. Another distinctive clay, sample 12, the slim, fingerlike bar, is richest in two micronutrients, zinc and manganese, but is weak in iron and copper. (1973: 177)

Hunter concluded that in view of the deficiencies inherent in the diets of many Africans, clays supplement those diets with a wide range of minerals important for pregnant women.

Moving to the other side of the world, Hunter and De Kleine (1984) also evaluated the nutritional properties of clay used to make tablets sold at religious shrines in Belize. Samples were tested for 13 minerals, and the conditions of digestion were simulated to determine the bioavailability of those minerals. Although the tests showed great variation in the amount of minerals available, in all cases the mineral content (and mineral availability) was sufficiently strong to suggest that the clay did indeed provide nutritional supplementation. A daily dose of from 1 to 6 tablets delivered 11 minerals in varying degrees, depending on the clay sampled. For example, 1 tablet yielded 73 milligrams of calcium and 68 micrograms of copper, among other minerals. A sample dose of 6 tablets provided 10 milligrams of iron, more than 200 micrograms each of nickel and manganese, and 12 micrograms of cobalt. When compared with U.S. Recommended Dietary Allowances and recommended supplements during pregnancy, a single tablet provided about 9 percent of the recommended iron and calcium.

Yet some types of clay are suspected of having properties that inhibit the absorption of minerals from the gastrointestinal tract, and there is no consistent agreement that iron from clay is useful in correcting anemia. Still, Hunter (1973), De Kleine (Hunter and De Kleine 1984), and David L. Browman and James N. Gundersen (1993) have asserted that clays can serve as a culturally acceptable and nutritionally functional source of mineral supplementation within some nonindustrialized societies.

In addition, there is an economic factor. Vermeer (1966, 1971) and Hunter (1973) have both docu-

mented how clay is incorporated into trade practices and how its processing has been the basis for local cottage industries.

Cultural Diffusion: Case Studies

Perhaps the best approach to understanding the phenomenon of geophagy involves the merging of nutritional, cultural, and economic explanations. In other words, from an African standpoint it might be viewed as a physiologically based adaptive behavior supported by religious and other cultural beliefs and institutionalized within the local economy.

Moreover, many have noted the diffusion of geophagy via the slave trade with the result that the practice was frequently observed among enslaved Africans in the West Indies, South America, and the southern United States. In addition, forms of geophagy continue to be documented among African-Americans in the southern United States and among those who migrated to northern cities and took the practice with them. It was common to have dirt or clay from a particular site in the South sent north by relatives. However, more recently, amylophagia (consumption of laundry starch) seems to have replaced clay eating, and geophagy generally seems to have declined significantly in recent decades.

During the last century, *cachexia africana* was of concern to physicians and plantation owners, who viewed it as an important cause of death among enslaved Africans. The condition they described seems similar to the syndrome of chlorosis we have already described. Both chlorosis and *cachexia africana* were characterized by a seemingly uncontrollable desire for the substances that are eaten. Although the reasons advanced for the aberrant appetites of young white women and African slaves are different, in both cases questions of cause and effect arise. David Mason (1833) was one of the first to suggest that obsessive dirt eating or *atrophia a ventriculo* (stomach atrophy), as he termed it, was actually a consequence of disease, and not its cause. He wrote:

The train of symptoms that progressively arise from atrophy of the stomach and dirt-eating are indigestion and emaciation; a bloated countenance; a dirty-yellow tinge in the cellular tissue of the eyelids; paleness of the lips and ends of the fingers; whiteness of the tongue; great indolence, with an utter aversion to the most ordinary exertion; palpitation of the heart; difficult, or rather frequent and oppressed, respiration, even during moderate exercise, which never fails to induce a rapid pulse; habitual coldness of the skin; and occasional giddiness of the head, attended with a disposition to faint, sometimes causing a state of stupor. (1833: 292)

In addition, he noted a change in the color and density of blood, the appearance of skin ulcers, polypi in

the heart, and pathologies of the liver and gallbladder. Mason also reported that the obsession with nonfood items extended to eating "cloth, both linen and woolen ..." (1833: 291).

Mason and later authors (Laufer 1930; Anell and Lagercrantz 1958) discussed a variety of conflicting explanations for the disease. Plantation owners often considered it either a kind of addiction or a means of escaping work, or both, and punished geophagy with confinement, beatings, and the use of metal mouthlocks. But others believed that homesickness and depression or abusive conditions led to the consumption of dirt; Berthold Laufer (1930) and Bengt Anell and Sture Lagercrantz (1958) mention the conviction, held by some, that geophagy was a deliberate, slow form of suicide among the enslaved Africans. Mason, by contrast, foreshadowed the current general debate about iron deficiency and pica in his belief that dirt eating was "rather a consequence than a cause of the disease" and suggested that the earth eaten contained "useful ingredients mixed up with much hurtful matter" and that "iron and alkalis are of great efficacy in this disease" (1833: 289, 292). He recommended exercise, proper diet, cleanliness, and proper clothing, as well as purely medical treatment with emetics, purgatives, and tonics. Unlike geophagy more generally, *cachexia africana* appeared to afflict both men and women in large numbers.

Hunter (1973) has noted similarities between geophagy in Africa and that practiced more recently by African-Americans in the United States, leading him along with others (see, for example, Vermeer and Frate 1975) to the view that geophagy arrived in the United States with enslaved Africans in a process of cultural diffusion. In both geographical locations the practice is found most often among women (especially pregnant women), and clays are gathered from special sites such as anthills, termitariums, or river banks and often referred to by the location (Forsyth and Benoit 1989). In Africa and in the southern United States, clays are shaped, baked, or sun-dried before eating and are claimed to have health-giving properties, as well as providing satisfaction. Reasons cited by southern women (black and white) reflect folk beliefs about the medicinal qualities of earth eating, while also indicating its character as an addiction:

I craves it. I eat dirt just the same way you would smoke a cigarette. I crave something sour like the taste of clay. It seems to settle my stomach. I know I shouldn't eat it. When I go up in Jasper County I get it, but can't find any good dirt here. This Biloxi dirt ain't no good, so I gets my sister in Birmingham to send it to me. I never heard of a man eating dirt. They not got the same taste a woman has. My mother eats it because she be's in the change and they say it will help her. When I was a child I was coming home from Sunday School and it had rained. I

could smell the dirt on the bank and started to eat it then. Have kept it up. I would eat more dirt than I do, but I have a hard time getting it. (quoted in Ferguson and Keaton 1950: 463)

In moving to non-Western case studies of pica, we can encounter one of the few of a psychological nature. This one concerns an “epidemic” of pica (geophagy) that took place in Aboriginal coastal towns of northern Australia during 1958 and 1959 (Eastwell 1979). As diarrhea was presumed to be the cause of the earth eating, white nurses in the area treated patients with placebo tablets and with Kaopectate. The “epidemic” was ended, but the cessation of geophagy produced a community-wide disorder of hypochondriasis, and another outbreak of geophagy occurred among Aboriginal women who were past the years of childbearing.

Eastwell (1979), who discussed the epidemic, believed that the hypochondriasis and the earth eating of postmenopausal women had deep sociocultural causes. He argued that within the aboriginal hunting-and-gathering mode of subsistence, geophagy was originally practiced for medicinal purposes, but this function disappeared after these small-scale societies were colonized by the British. As the traditional hunting and gathering came to an end, the practice of geophagy was transformed into an indigenous cultural statement of Aboriginal status. Thus, bereft of its traditional medicinal function, the practice of pica among Aboriginal women became not only a type of protest, reflecting gender status, but also constituted a psychological adjustment strategy for women, whose economic importance diminished during the shift from a nomadic to a sedentary existence.

In a separate example, involving cultural diffusion and social change, Hunter and De Kleine (1984) provided a case study demonstrating the interaction of cultural and nutritional aspects of geophagia in Belize, where the indigenous population was resettled by the Spaniards in Santiago de Esquipulas, a pre-Columbian town of economic and religious importance to the Maya. “Esquipulas was noted for its shrine, health-giving earth, and sulfurous springs; thus it served as a place of spiritual significance and healing activity” (Hunter and De Kleine 1984: 157). The Spaniards built a chapel at the site to house a crucifix carved of balsam and orangewood. The crucifix darkened over time with the burning of incense and candles and became known as *Nuestro Señor de Esquipulas, el Cristo Negro*, or “the Black Christ.” The Black Christ was worshiped by the Indians and symbolized a cultural fusion of Christian and Mayan beliefs by bringing an “Indian saint” into what was a new religion for the Mayas. The Black Christ also became known for its miraculous cures, which focused attention on the healing properties of the spring and mud at the site.

In the 1700s, the site was formally recognized by

the Catholic Church, and a sanctuary for the Black Christ was constructed. The shrine continues to be of religious importance; pilgrims visit it, particularly on January 15, the Day of Esquipulas, and during Lent. Prior to the institution of border regulations requiring passports and visas, an estimated 100,000 pilgrims annually visited the shrine to be cured of ailments including “leprosy, blindness, muteness, insanity, paralysis, rabies, yellow fever, malaria, tetanus, and hemorrhages” (Hunter and De Kleine 1984: 158).

Many of the alleged cures were credited to the *tierra santa* (holy clay) at the site, which is believed to have health-giving properties and is blessed by the Roman Catholic Church. *Tierra santa* was (and is) sold at the shrine in the form of clay pressed into small cakes, stamped with images of the Virgin Mary, the Black Christ, and other saints. The clay tablets or *benditos* (blessed ones) (either eaten or dissolved in water and drunk) are believed to alleviate diseases of the stomach, heart, and eyes, to ease menstrual difficulties, and to facilitate pregnancy and childbirth.

As the cult of the Black Christ and the reputation for cures spread throughout Central America, new shrines were built that also became associated with curative, blessed earth. Indeed, by the end of the eighteenth century there were shrines in at least 40 towns where supposedly curative earths were available for consumption.

Pagophagia and Amylophagia

Within Western industrialized societies, recent medical literature on pica is dominated not so much by geophagy as by pagophagia and amylophagia. The term “pagophagia,” first used in 1969 by Charles Colman, a U.S. Air Force physician, refers to the compulsive consumption of ice and other frozen substances. Some, however, do not view pagophagia as a form of pica because ice (and ice water) consumption can be a positive measure in controlling body weight and addictions such as the use of tobacco. In addition, of course, chewing ice is more socially acceptable within industrialized countries than eating dirt. Like geophagy, pagophagia is strongly associated by many researchers with iron and other mineral deficiencies.

Amylophagia, by contrast, is the eating of laundry starch, which is associated almost exclusively with women. It was first observed in rural areas of the southeastern United States, where it was thought by some to have replaced dirt eating. As we have already noted, Hunter described a process of cultural diffusion and change occurring as geophagy from Africa was brought first to the southern United States:

Next came the northward migration of blacks to the urban ghettos of Cleveland, Chicago, New York, and Detroit. Such migrants ask their southern relatives to mail them boxes of clay for consumption during pregnancy. At this

state, however, the forces of culture conflict come to the fore: lack of local clay in the concrete jungles of the North, pressures of poverty, and stress on kinship ties with the South lead to the consumption of laundry starch replacing traditional geophagy. But micronutrient minerals are totally lacking in the starch. Calories apart, nutritional inputs are zero; gastric irritation is caused. A cultural practice is now divorced from nutritional empiricism; cultural adjustment to socioenvironmental change has broken down, and atrophy and decay are the result. (1973: 193)

Perhaps significantly, a similarity between dirt and laundry starch in texture (although not in taste) has been noted by investigators. The reasons cited by women for consuming laundry starch include the alleviation of nausea and vomiting associated with pregnancy and various folk beliefs, found largely among African-Americans, that consuming starch during pregnancy helps the baby "slide out" during delivery, promotes a healthy baby, or a whiter (or darker) baby (O'Rourke et al. 1967).

Like pagophagia and geophagia, as well as other forms of pica, amylophagia has also been associated with iron deficiency. Deleterious consequences include impacted bowels and intestinal obstructions.

Pica during Pregnancy

Pica has been associated since classical times with pregnant women. Until the twentieth century, pregnancy was commonly believed to cause mental instability - manifested, for example, in unusual food cravings. More recent studies of food preferences during pregnancy, however, report that changes in these, as well as the onset of specific cravings, are not universal phenomena.

Much of the research on pica among pregnant women in the United States has focused on those living in rural areas. The prevalence of pica among women considered at risk seems to have declined by about half between 1950 and 1970, but it has remained fairly constant from 1970 to the present. Nonetheless, it was estimated that pica is practiced by about one-fifth of pregnant women in the United States who are considered at "high risk" for this behavior. "High-risk" factors include being African-American, living in a rural area, having a family history of pica, and having practiced pica during childhood (Horner et al. 1991). Pregnant black women are over four times more likely than their white counterparts to engage in pica behavior. Additionally, pregnant women living in rural areas are more than twice as prone to pica as those living in urban areas.

Although some investigators have found no significant association between age and pica among pregnant women (Dunston 1961; Butler 1982), others

have observed that pregnant women who practice pica tend to be relatively older than those who do not. It is interesting to note that women who report consuming clay tend to be older than those who report consuming starch (Vermeer and Frate 1979). One study, reanalyzing data from previous research, indicated that pregnant women who did practice pica were six times as likely to have a history of childhood pica than pregnant women who did not. Women who practice pica during pregnancy are also more likely to report pica behavior among family members, particularly their mothers and grandmothers (Lackey 1978). Little evidence of pica among white and upper-income women may reflect a lack of research among these populations (Keith, Brown, and Rosenberg 1970; Horner et al. 1991).

Among pregnant women in the United States, the three forms of pica that occur most frequently are geophagia, amylophagia, and pagophagia (Horner et al. 1991). Although, as we have noted, some researchers believe that as African-American women migrated to northern urban areas, laundry starch became a substitute for the more traditional clay eaten in the South (Keith et al. 1970), other research indicates that consumption preferences themselves might be changing, with younger women preferring starch over clay. In one study of rural women in North Carolina, participants indicated a preference for starch, even though their mothers had consumed both clay and starch (Mansfield 1977). Explanations of pica during pregnancy, like those of pica in general, range from the psychological through the cultural, to the nutritional (Horner et al. 1991; Edwards et al. 1994).

A recent study of eating habits and disorders during pregnancy mentions the case of a woman who, at 32 weeks of pregnancy, developed a craving for coal, reporting she found it "irresistibly inviting" (Fairburn, Stein, and Jones 1992: 668). Two other participants in the study developed a taste for eating vegetables while still frozen, which indicates something of the difficulty involved in determining pica incidence. The consumption of frozen vegetables, although not defined as pica by these researchers, would surely be considered a type of pagophagia by others.

In terms of medical consequences, pica has been related to anemia and toxemia among pregnant women and newborn infants (Horner et al. 1991). In some cases, pica reportedly contributed to dysfunctional labor (through impacted bowels) and maternal death (Horner et al. 1991). Pica during pregnancy has also been associated with a "poor" functional status of fetuses and infants, perinatal mortality, and low birth weight.

The authors of a report on pica in the form of baking-powder consumption that caused toxemia during pregnancy have pointed out that previous investigators discovered a significant correlation between toxemia and geophagia, but not between toxemia and

amylophagia (Barton, Riely, and Sibai 1992). The case involved a 23-year-old black woman with anemia and hypokalemia who admitted to a one-and-a-half-year history of consuming up to 7 ounces of Calumet baking powder daily. The baking powder was considered a family remedy for gas discomfort. Ingestion of baking powder, comprised of 30 percent sodium bicarbonate with cornstarch, sodium aluminum sulfate, calcium acid phosphate, and calcium sulfate, is known to increase blood pressure. In this case, liver dysfunction and hypokalemia also resulted.

The psychological aspects of pica among pregnant women are similar to those of other pica practitioners. In addition to reporting a craving for the ingested substance, pregnant women exhibiting pica commonly say that they feel anxious when the substance is unavailable yet experience a sense of considerable satisfaction during and after eating the substance (Horner et al. 1991).

Pica and Iron Deficiency

Although no definitive connection has been established between pica and nutritional deficiencies, many have consistently linked pica with iron deficiency and its consequent anemia. Indeed, some have estimated that upward of 50 percent of patients with iron-deficiency anemia practice pica (Coltman 1969; Crosby 1976). It is interesting to note that the correlation of pica with anemia dates back to medieval times, and that iron therapy was prescribed as a cure even then (Keith et al. 1970).

As pointed out previously, pica behavior during pregnancy has also been strongly associated with iron-deficiency and iron deficient anemia. What remains unresolved is a problem of cause and effect. As Dennis F. Moore and David A. Sears wrote: "Some authors have suggested that the habit may induce iron deficiency by replacing dietary iron sources or inhibiting the absorption of iron. However, considerable evidence suggests that iron deficiency is usually the primary event and pica a consequence" (1994: 390). Although some insist that ingested starch inhibits iron absorption, Kenneth Talkington and colleagues (1970) have reported that this is not the case. These authors concluded that iron deficiency and anemia result from amylophagia only when laundry starch replaces nutritional substances in the diet.

Turning to clay ingestion, studies have found that its effect on iron absorption varies and depends upon the type of clay ingested. Some clays impair iron absorption, whereas others contain large amounts of iron. However, as already mentioned, there is no consistent agreement that iron from clay is useful in correcting anemia (Coltman 1969; Keith et al. 1970; Crosby 1976).

Coltman (1969), who first used the term pagophagia, was also one of the first to link the practice with iron deficiency. Indeed, he reported that the compulsive

consumption of ice could be stopped within one or two weeks with iron treatment, even in instances where iron supplementation was not sufficient to correct iron-deficiency anemia. This dovetails with the work of William H. Crosby (1976), who has noted that although ice neither displaces other dietary calories nor impairs iron absorption, it is still the case that pagophagia is diminished when treated with iron supplements. Moreover, other cases of pica involving unusual ingested substances (e.g., toothpicks, dust from venetian blinds, and cigarette ashes) also respond positively to iron supplements (Moore and Sears 1994).

Perhaps even more powerful support for iron deficiency as a cause of pica comes from findings that intramuscular injections of iron diminish the habit of pica in children. But there is also evidence that intramuscular injections of a saline solution have the same effect, suggesting that the additional attention paid to children with pica behavior may help to reverse the condition (Keith et al. 1970).

Countering this theory, however, are two cases of childhood pica in which parental attention was apparently not a factor. One involved a 6-year-old boy with a 2-year history of ingesting large amounts of foam rubber, whereas the second case was that of a 2-year-old boy with a 6-month history of eating plastic and rubber items. Both of these cases of pica behavior were resolved through the administration of iron supplements, even though there was no increase in parental attentiveness (Arbiter and Black 1991).

In 1970, Louis Keith, E. R. Brown, and L. Rosenberg summarized the medical questions surrounding pica that required further investigation and clarification. These were:

1. Is iron-deficiency anemia a direct adverse consequence of pica?
2. Is iron-deficiency anemia an indirect result of nutritional replacement by unnatural substances, allaying the appetite for nutritional foods by filling?
3. Are so-called cures of the habit of pica among children the result of increased attention or of injections of iron or saline solution?
4. If those "cures" among children are the result of therapy, should the therapy consist of iron injections, saline injections, or an adequate diet high in iron content?
5. Is the mechanism of pica among children different from that among adults, especially pregnant women? Would injections of iron reverse the habit in pregnant women?
6. Does the coexistence of amylophagia and anemia adversely affect the pregnant woman, or are these two separate and distinct unrelated concomitant adverse conditions? (1970: 630)

As we have seen, although almost three decades have elapsed, there is still no consensus within the medical community regarding the answers to these questions. Instead, they are still being asked.

The physiological mechanism linking iron deficiency and pica behavior is not known. As Crosby noted, "Somewhere in our emotional circuits iron deficiency can sometimes cross the wires" (1976: 342). Somewhat more scientifically, it has been suggested that pica cravings are generated by a functional disorder of the hypothalamus, which is sensitive to changes in iron levels (Castiglia 1993).

As for pagophagia, Mary Elks (1994) has made two observations. She noted that even in industrialized nations, both geography and culture play a role in determining how pagophagia is viewed. For example, in England and other European countries compulsive ice eating is considered pathological behavior, perhaps indicative of disease. However, in the warmer climate of the southern United States, ice eating may be regarded as normal. In addition, she reported a case involving an entire family practicing pagophagia, including a 14-month-old girl for whom, Elks believed, ice consumption should not be assumed to be a learned behavior. She suggested that in some cases, familial or heritable factors cause pagophagia that is independent of other types of pica and probably not correlated with iron deficiency.

Psychological Explanations of Pica

Pica, as we have observed, is included as an eating disorder (along with anorexia nervosa, bulimia, and rumination in infancy) in the classification systems of the DSM-III-R of the American Psychiatric Association (APA) (1987) and the ICD-10 of the World Health Organization (1992). Pica is defined by the APA as the repeated consumption of nonnutritive substances for a period of at least one month, when the behavior is not attributable to another mental disorder.

It is interesting to note that psychological literature on eating disorders discusses pica in early childhood as a risk factor for bulimia in adolescence (Marchi and Cohen 1990). Moreover, some aspects of pica, such as excessive consumption of ice, ice water, lemon juice, and vinegar, are linked with anorexia nervosa (Parry-Jones 1992; Parry-Jones and Parry-Jones 1994). Pica has also been associated with rumination in children and persons with mental retardation; in such cases, the behavior may be interpreted as a regressive behavior reflecting oral needs that have not been met (Feldman 1986). Poor feeding and weaning practices are more frequently observed in children with pica than in those without (Singhi, Singhi, and Adwani 1981). In addition, comparisons of children who have iron-deficiency anemia and who practice pica and children with anemia who do not

practice pica show that the former score higher on measures of stress, including that caused by maternal deprivation, child abuse, and parental separation (Singhi et al. 1981).

It is important to understand that psychological explanations themselves tend to reflect cultural beliefs in the Western industrialized world, in contrast to religious and spiritual beliefs of cultures in other areas. To view all types of pica behavior as pathology risks a failure to recognize other important issues, such as cultural variation in food preference, indigenous medicinal and nutritional knowledge, and the very real question of the effects of nutritional deficiencies.

Discussion

Although explanations of pica vary with the type of substance consumed and with the cultural context, there are several consistent themes in the literature. One is that whereas men do practice pica, the behavior is most frequently associated with women and children. Second, regardless of cultural context and whether pica is considered acceptable among adults, there seems to be a uniform concern about the practice of pica by children.

A third has to do with similarities shared by the substances most frequently consumed. They tend to be brittle, dry, and crunchy. Moreover, the smell of the clay and soil is cited as important in a variety of cultural contexts, as is the location from which the clay or earth is obtained. These sites are frequently the homes of living things such as termites or crayfish. There is also a pattern, historically and geographically, of consumption of earth or clay from sites of special significance. Such sites may be religious, as in Latin America and Africa, locally distinctive, as in the southern United States, or they may be places of burial, as in Asia, Africa, and Europe. Finally, there is a cross-cultural consistency in the debate concerning the relationship of pica to iron and other mineral deficiencies, and in the debate over whether the practice is psychologically or physiologically based.

One question that remains unexplored (in fact, it is barely mentioned in the literature) is why the definitions and syndromes associated with pica are not extended to the many practices in which men rather than women more typically engage, such as chewing tobacco and cigars or pipe stems, using snuff, and chewing toothpicks, betel nut, and chewing gum. While these habits may not generally involve consumption, they are not so dissimilar from the practices of pagophagia. Certainly, it appears that the relationship between the forms of pica we have discussed and other cravings and sources of oral satisfaction is an area in need of further investigation.

Margaret J. Weinberger

Bibliography

- American Psychiatric Association. 1987. *Diagnostic and statistical manual of mental disorders (DSM-III-R)*. Washington, D.C.
- Anell, Bengt, and Sture Lagercrantz. 1958. *Geophagical customs*. Studia ethnographica Upsaliensia 17, Uppsala, Sweden.
- Arbiter, E. A., and Dora Black. 1991. Pica and iron-deficiency anaemia. *Child: Care, Health, and Development* 17: 231–4.
- Barton, John R., Caroline A. Riely, and Baha M. Sibai. 1992. Baking powder pica mimicking preeclampsia. *American Journal of Obstetrics and Gynecology* 167: 98–9.
- Bell, Kenneth E., and David M. Stein. 1992. Behavioral treatments for pica: A review of empirical studies. *International Journal of Eating Disorders* 11: 377–89.
- Browman, David L., and James N. Gundersen. 1993. Altiplano comestible earths: Prehistoric and historic geophagy of highland Peru and Bolivia. *Geoarchaeology* 8: 413–25.
- Butler, P. M. 1982. Pica practices as an influence on iron deficiency anemia. Ph.D. thesis, East Carolina University, Greenville, N.C.
- Castiglia, Patricia T. 1993. Pica. *Journal of Pediatric Health-Care* 7: 174–5.
- Coltman, Charles A. 1969. Pagophagia and iron lack. *Journal of the American Medical Association* 207: 513–14.
- Cooper, Marcia C. 1957. *Pica*. Springfield, Ill.
- Crosby, William H. 1971. Food pica and iron deficiency. *Archives of Internal Medicine* 127: 960–1.
1976. Pica: A compulsion caused by iron deficiency. *British Journal of Haematology* 34: 341–2.
- Danford, D. E. 1982. Pica and nutrition. *Annual Review of Nutrition* 2: 303–22.
- Dannenfeldt, K. H. 1984. The introduction of a new sixteenth-century drug: *terra silesiaca*. *Medical History* 28: 174–88.
- Dunston, B. N. 1961. Pica, hemoglobin and prematurity and perinatal mortality. Ph.D. dissertation, New York University School of Education, New York.
- Eastwell, Harry D. 1979. A pica epidemic: A price for sedentarism among Australian ex-hunter-gatherers. *Psychiatry* 42: 264–73.
- Edwards, C. H., A. A. Johnson, E. M. Knight, et al. 1994. Pica in an urban environment. *Journal of Nutrition* 124 (Supplement): 954S–62S.
- Elks, Mary. 1994. Familial pagophagia. *Southern Medical Journal* 87: 963–4.
- Fairburn, Christopher G., Alan Stein, and Rosemary Jones. 1992. Eating habits and eating disorders during pregnancy. *Psychosomatic Medicine* 54: 665–72.
- Feldman, Marc D. 1986. Pica: Current perspectives. *Psychosomatics* 27: 519–23.
- Ferguson, James H., and Alice Glenn Keaton. 1950. Studies of the diets of pregnant women in Mississippi: The ingestion of clay and laundry starch. *Louisiana State Medical Society Journal* 102: 460–3.
- Forsyth, Craig J., and Genevieve M. Benoit. 1989. Rare, ole, dirty snacks: Some research notes on dirt eating. *Deviant Behavior* 10: 61–8.
- Halsted, J. A. 1968. Geophagia in man: Its nature and nutritional effects. *American Journal of Clinical Nutrition* 21: 1384–93.
- Higgins, Brian. 1993. Pica. In *The Cambridge world history of human disease*, ed. Kenneth F. Kiple, 297–332. Cambridge and New York.
- Hooper, R. 1811. *Quincy's lexicon-medicum. A new medical dictionary*. London.
- Horner, R. D., C. J. Lackey, K. Kolasa, and K. Warren. 1991. Pica practices of pregnant women. *Journal of the American Dietetic Association* 91: 34–8.
- Horst, Oscar H. 1990. Arcilla geofagica en America. *Meso-america-antigua* 19: 169–76.
- Hunter, John M. 1973. Geophagy in Africa and the United States: A culture-nutrition hypothesis. *Geographical Review* 63: 170–95.
1984. Insect clay geophagy in Sierra Leone. *Journal of Cultural Geography* 4: 2–13.
- Hunter, John M., and Renate De Kleine. 1984. Geophagy in Central America. *Geographical Review* 74: 157–69.
- Johns, Timothy, and Martin Duquette. 1991. Detoxification and mineral supplementation as functions of geophagy. *American Journal of Clinical Nutrition* 53: 448–56.
- Karp, Joyce Gerdis, Laura Whitman, and Antonio Convit. 1991. Intentional ingestion of foreign objects by male prison inmates. *Hospital and Community Psychiatry* 42: 533–5.
- Keith, L., E. R. Brown, and C. Rosenberg. 1970. Pica: The unfinished story; background; correlations with anemia and pregnancy. *Perspectives in Biology and Medicine* 13: 626–32.
- Kiple, Kenneth. 1984. *The Caribbean slave: A biological history*. Cambridge.
1988. ed. *The African exchange: Toward a biological history of black people*. Durham, N.C.
- Kiple, Kenneth, and V. Himmelsteib King. 1981. *Another dimension to the black diaspora: Diet, disease and racism*. Cambridge.
- Lackey, C. J. 1978. Pica – a nutritional anthropology concern. In *The anthropology of health*, ed. E. E. Bauwens, 121–9. St. Louis.
- Laufer, Berthold. 1930. Geophagy. In *Anthropological series*, Field Museum of Natural History, Vol. 18 (2): 99–198, Chicago.
- Loudon, I. S. L. 1980. Chlorosis, anaemia and anorexia nervosa. *British Medical Journal* 281: 1669–75.
- Mansfield, C. 1977. *Investigation of pica in Pitt County, North Carolina*. Greenville, N.C.
- Marchi, M., and P. Cohen. 1990. Early childhood eating behaviours and adolescent eating disorders. *Journal of the American Academy of Child and Adolescent Psychiatry* 29: 112–17.
- Mason, David. 1833. On atrophica ventriculo (mal d'estomac), or dirt eating. *Edinburgh Medical and Surgical Journal* 39: 289–96.
- Melville, Bendley, and Valery Francis. 1992. Dietary habits and superstitions of rural Jamaican women during pregnancy. *European Journal of Clinical Nutrition* 46: 373–4.
- Moore, Dennis F., and David A. Sears. 1994. Pica, iron deficiency, and the medical history. *The American Journal of Medicine* 97: 390–3.
- O'Brien, Robert. 1985. *The encyclopedia of the South*. New York.
- O'Rourke, D. E., J. G. Quinn, J. O. Nicholson, and H. H. Gibson. 1967. Geophagia during pregnancy. *Obstetrics and Gynecology* 29: 581–4.
- Parry-Jones, Brenda. 1992. Pagophagia or compulsive ice consumption: A historical perspective. *Psychological Medicine* 22: 561–71.
- Parry-Jones, Brenda, and William L. Parry-Jones. 1992. Pica: Symptom or eating disorder? A historical assessment. *British Journal of Psychiatry* 160: 341–54.

- Parry-Jones, William L., and Brenda Parry-Jones. 1994. Implications of historical evidence for the classification of eating disorders: A dimension overlooked in DSM-III-R and ICD-10. *British Journal of Psychiatry* 165: 287–92.
- Pope, Janet F., Jean D. Skinner, and Betty R. Carruth. 1992. Cravings and aversions of pregnant adolescents. *Journal of the American Dietetic Association* 92: 1479–82.
- Prince, Isolde. 1989. Pica and geophagia in cross-cultural perspective. *Transcultural Psychiatric Research Review* 26: 167–97.
- Rehm, M. D., and Philip A. DeSimone. 1991. A proposed mechanism for cardboard-induced iron-deficiency anemia. *The American Journal of Medicine* 90: 768–9.
- Reid, Russell M. 1992. Cultural and medical perspectives on geophagia. *Medical Anthropology* 13: 337–51.
- Singhi, S., P. Singhi, and G. B. Adwani. 1981. Role of psychosocial stress in the cause of pica. *Clinical Pediatrics* 20: 783–5.
- Talkington, Kenneth, Norman F. Gant, Daniel E. Scott, and Jack A. Pritchard. 1970. Effects of ingestion of starch and some clays on iron absorption. *American Journal of Obstetrics and Gynecology* 108: 262–7.
- Vermeer, Donald E. 1966. Geophagy among the Tiv of Nigeria. *Association of American Geographers* 56: 197–204.
1971. Geophagy among the Ewe of Ghana. *Ethnology* 10: 56–72.
- Vermeer, Donald E., and Dennis A. Frate. 1975. Geophagy in a Mississippi county. *Annals of the Association of American Geographers* 65: 414–24.
1979. Geophagia in rural Mississippi: Environmental and cultural contexts and nutritional implications. *American Journal of Clinical Nutrition* 32: 2129–35.
- World Health Organization. 1992. *The ICD-10 classification of mental and behavioural disorders*. Geneva.

IV.D.7 Protein–Energy Malnutrition

Protein–energy malnutrition (PEM) is the current term for a group of nutritional diseases related to dietary protein and energy (calorie) intake. These diseases are most frequently seen in infants and young children in developing countries but may be a feature of famine or the result of illness for people of all ages throughout the world. Research during the twentieth century has considerably clarified the causes and manifestations of what are now known as dietary-related effects on the growing or mature individual. PEM includes conditions known in the medical world as kwashiorkor, marasmus, and growth retardation in children. Related to PEM are pellagra, starvation, and protein malnutrition. Infection, debilitating disease, and surgical procedures are frequently complicated by PEM. It is, therefore, a factor of importance in determining morbidity and mortality and has to be taken into account by health-care personnel at all levels.

Historical Concepts of Protein and Energy

Early reports of what may have been PEM lack the clinical, pathological, and biochemical details that make identification certain. The history of PEM is thus confined to the nineteenth and twentieth centuries, and it is only in the last 50 years that clarification of the various forms that PEM can manifest has emerged.

In his book *Protein and Energy*, Kenneth J. Carpenter has provided a detailed survey of nutritional science as it was known in the period from 1614 to 1893 (Carpenter 1994: 1–99). Of particular interest in relation to later discoveries is that the first “balance studies” were carried out by Italian scientist S. Santorio in 1614. He weighed his food and drink as well as his excreta (urine and feces) and measured changes in his own weight. There was an unexplained daily disappearance of 5 pounds of material that he attributed to a breakdown of body tissue that was then secreted through the skin as insensible perspiration; the losses were made good by the nourishment ingested. This was only a more quantitative restatement of Galen’s view in the second century that “[o]ur bodies are dissipated by the transpiration that takes place through the pores in our skins that are invisible to us; therefore we need food in a quantity proportionate to the quantity transpired” (Carpenter 1994: 1).

Anton Lavoisier’s work in the late eighteenth century (1770 to 1790) made him the “Father of Nutritional Science.” His first contribution was the recognition of the distinction between compounds that could change their character and simple substances or elements (e.g., carbon, hydrogen, nitrogen, oxygen, and others). His second contribution was an understanding that combustion and respiration involved similar processes of oxidation that could explain the phenomenon of “animal heat.”

Among those who followed Lavoisier was Jean-Baptiste Boussingault, who published the first table of the nitrogen content of foods in 1836. The protein radical was discovered just two years later by Gerrit Mulder and was considered to be the essential ingredient for both body building and physical activity. By the end of the nineteenth century, however, it was realized that protein was not the main or obligatory source of energy and that it is the oxidation of carbohydrates and fatty acids on which we rely for continued physical work.

Definition and Nomenclature

Description of the various syndromes that we now include in PEM began around 1850, and in the subsequent hundred years, much confusion in terminology arose. However, E. Kerpel-Fronius, a well-known Hungarian pediatric investigator who bridged both the pre- and post-1950 eras, clarified the old and new nomenclature, which has made it possible to identify

references in early case reports (Kerpel-Fronius 1983: 30–4). He showed that the contradictions in terminology were rooted in regional differences, age of weaning, local foods, and prevalence of infections, and he classified the various types of malnutrition we now call PEM in the following way:

1. Hypoalbuminemic forms (low serum proteins)
 - i. Edematous
 - a. Kwashiorkor
 - b. Mehlnárschaden
 - ii. Hypoalbuminemic forms without edema
2. Dry forms without hypoalbuminemia
 - i. Underweight (dystrophic infants)
 - a. Stationary stage
 - b. Repairing stage with retardation in height (stunted infants)
 - ii. Marasmus (atrophy)
 - a. Moderately severe
 - b. Severe forms
 - iii. Severest form in young infants (athrepsia or decomposition in the classic texts of pediatrics)

Reference to this classification assists considerably in identifying in older literature the various forms of what is now known as PEM. A current definition of PEM is the Wellcome classification shown in Table IV.D.7.1, which refers to children (Wellcome Trust Working Party 1970), whereas a broad current understanding of the nutritional basis of PEM is as follows:

1. Protein quantity and/or quality (amino acid pattern of the protein) intake that is below the minimal requirements for growth and health, with or without an energy intake that is less than energy expenditure on muscle activity, heat production, growth, and other energy requirements.
2. Excessive loss of protein and energy in diarrhea and acute and chronic diseases. (Waterlow 1992: 152–8)

PEM can, of course, be complicated by mineral deficiencies (e.g., of sodium, potassium, calcium, phosphorus, or iron), by deficiencies of trace elements (e.g., zinc or chromium), and by vitamin deficiencies (e.g., of vitamins A, D, C, or K). Much of this understanding has come about through intensive

worldwide research during the last 50 years, which is discussed in the remainder of this chapter. We begin with marasmus.

Marasmus 1850 to 1950

“Marasmus” means “wasting away of the body” (from the Greek *marasmos*) and is a term applied mainly to infants and children (the term “wasting” is employed for similarly afflicted adults) (Figure IV.D.7.1). It occurs when the diet is grossly deficient in energy. Such a diet also necessarily fails to meet protein requirements. Marasmus may become manifest in wholly breast-fed infants when the milk is quantitatively insufficient, but more frequently it occurs after



Figure IV.D.7.1. Marasmus. Age about 7 months grossly underweight. No skin lesions and no edema. (Photograph courtesy of Groote Schuur Hospital, Cape Town, South Africa.)

Table IV.D.7.1. *The Wellcome classification of PEM*

Weight (percentage of standard)	Edema	
	Present	Absent
60–80	Kwashiorkor	Underweight
60 or less	Marasmic Kwashiorkor	Marasmus

Source: Wellcome Trust Working Party: 1970.

early weaning to diluted or low-energy bottle feeds or cereal paps. In the age group from 1 to 5 years, marasmus occurs when food of any kind is in short supply, as in conditions of war, civil unrest, famine, extreme poverty, or just lack of care for the child. Often, it is produced by starvation that occurs during treatment of diarrhea or other infections, but it can also stem from severe weight loss brought on by chronic pyogenic disease, tuberculosis, syphilis, AIDS, and tropical infestations.

The presenting symptoms are failure to thrive, irritable crying, or apathy. Diarrhea is frequent, and the child has a shrunken or wizened appearance, even though it is ravenously hungry. The degree of underweight for age is extreme, and the muscles are weak and atrophic. Currently, marasmus is diagnosed when the weight is 60 percent or less of the 50th percentile of the National Center for Health Statistics (NCHS) standards (Table IV.D.7.1). However, based on prewar literature, Kerpel-Fronius wrote that if the degree of wasting reaches 35 to 40 percent of the average body weight, recovery is impossible (Kerpel-Fronius 1947).

An interesting description of the clinical signs of marasmus was given in a speech by Charles Dickens at a fund-raising dinner for the Hospital for Sick Children on February 9, 1858. After discussing the tens of thousands of children who were dying because of poverty and sickness, he described a tour of the old town of Edinburgh, where lived some of the city's worst-lodged inhabitants. In one of the most wretched dwellings, "there lay, in an old egg box, which the mother had begged from a shop, a little feeble, wasted, wan, sick child. With his little wasted face and his little hot, worn hands folded over on his breast, and his little bright, attentive eyes looking steadily at us" (Dickens 1956:607).

Prevalence

Exact figures of the prevalence of marasmus are difficult to obtain because of the confusion of nomenclature and diagnosis. But marasmus must have been a leading cause of morbidity and mortality among infants and preschool children during the latter half of the nineteenth century and the early part of the twentieth century – and in Europe and North America as well as the rest of the world. A large percentage of the victims were illegitimate foundlings, who represented from 15 to 45 percent of all newborn infants in most European capitals. In the poor hygienic conditions of the "foundlings' homes," death claimed between 30 and 90 percent during the first year of life.

Similar rates of mortality occurred in North American institutions. H. D. Chapin (1915), examining 11 foundling institutions in New York, discovered a death rate of 422 per 1,000 for children under 2 years of age – compared with a community-wide figure of only 87 per 1,000. In 1921, Oscar M. Schloss,

the author of the annual report of the Infants' Hospital in Boston, called attention to "the frequent relationships of both malnutrition and acute infections to infant deaths" and stressed their importance as fields of research (Smith 1983: 138). Marasmus and atrophy do not appear as diagnoses in lists of patient admissions from the period 1883 to 1913, but figures for debility, dysentery, and diarrhea are high. In 1882, more than one-third of admissions were for "debility."

With the two world wars, marasmus was especially prevalent in besieged cities of Europe, such as Warsaw, Leningrad, and Budapest. But by 1950, and since then, marasmus resulting from poverty and sickness has become very rare in the so-called developed world. It remains, however, a nutritional problem in the developing countries of Africa, the Middle East, South and Central America, and Southeast Asia (Bellamy 1996:98).

Research

From 1850 to 1950, numerous medical research reports on marasmus appeared, and many of these have been outlined, at least in part, in recent publications (Kerpel-Fronius 1983; Hansen 1991). Early in the twentieth century, German and French pediatricians stressed that the fate of undernourished infants depended on whether they escaped infection: It was thought that a rapid decline in the weight curve, ending in severe marasmus, could seldom be caused just by semistarvation (Kerpel-Fronius 1983:64–8).

In a 1905 analysis of the bodies of marasmus victims and normal infants, a striking difference was discovered in fat content. Fat content in marasmic infants was only 1.7 percent of body weight, compared with 13.1 percent in normal babies. In 1920, body water in marasmic infants was found to be increased from 70 to 80 percent, and it was noted that the brain, kidneys, and skeleton were relatively preserved in marasmus.

Ten years earlier dextrose saline had been used for collapsed and dehydrated cases – including those of infantile atrophy, as recorded in an article about the activities of the Boston Floating Hospital (Beaven 1957). Losses of potassium, sodium, and chloride in diarrhea had been demonstrated by K. Schmidt in 1850, L. F. Meyer in 1910, L. Tobler in 1911, and I. Jundell in 1913 (Darrow 1946). In 1915, it was confirmed that increased water, fat, chloride, sodium, and potassium losses occurred in loose stools of infants, including those with marasmus (Holt, Courtney, and Fales 1915). The authors of this study pointed out that in therapy, potassium and magnesium were needed in addition to water and sodium, but it was not until 1946 (when it was proved that potassium could safely be added to intravenous fluids) that this advice was followed. This addition cut mortality from 32 to 6 percent (Darrow 1946; Govan and Darrow 1946).



Figure IV.D.7.2. Kwashiorkor. Age 10 months. Skin lesions and edema. (Photograph courtesy of Groote Schuur Hospital, Capetown, South Africa.)

Kwashiorkor to 1954

Marasmus was the first syndrome of what we now know as PEM to become the focus of wide interest, concern, and research, especially between 1850 and 1950. But from 1935 onward, kwashiorkor became another intensively studied and important nutritional disease. The term “kwashiorkor” was introduced by Cicely Williams, who wrote: “The name kwashiorkor indicates the disease the deposed baby gets when the next one is born, and is the local name in the Gold Coast for a nutritional disease of children, associated with a maize diet” (1935: 1151-2) (Figure IV.D.7.2). Williams explained (1973) that the word comes from the Ga language of Accra, Ghana, and J. C. Waterlow (1991) has identified at least 31 distinct vernacular names in tropical Africa. Other regions also have words that doubtless frequently mean kwashiorkor. Among the common names in English, for example, are nutritional dystrophy, infantile pellagra, nutritional edema, and wet marasmus. In Spanish-speaking countries, kwashiorkor was referred to as a multiple deficiency state – *distrofia pluricarenal*.

Before naming the disease, Williams had published

a paper giving its clinical description (Williams 1933). Afflicted infants were of preschool age (1 to 4), and their diet generally involved breast feeding with supplementary feeds of maize paps low in protein content. On examination, there was edema, wasting diarrhea, sores of mucous membranes, desquamation of the skin on the legs and forearms, and a fatty liver. The disease was uniformly fatal unless treated. This description provided by Williams fits perfectly with our current clinical definition of kwashiorkor, except that since 1933, retardation of growth has been more emphasized and there is now detailed knowledge of the changes in function of various systems and organs (Hansen and Pettifor 1991).

In a 1952 World Health Organization (WHO) report, the name kwashiorkor was applied to the syndrome in Africa, and the relationship of the disease to a low-protein diet was firmly established (Brock and Autret 1952). This study and report had been initiated by the Joint FAO (Food and Agriculture Organization)/WHO Expert Committee on Nutrition at its first session in Geneva in October 1949. The committee had found that one of the most widespread nutritional disorders in tropical and subtropical areas was an ill-defined syndrome known by various names such as kwashiorkor, malignant malnutrition, polydeficiency disease, and so forth. It was resolved that WHO conduct an inquiry into the various features of kwashiorkor, and, subsequently, the FAO was asked to cooperate. J. F. Brock (WHO) and M. Autret (FAO) traveled extensively throughout Africa over a period of two months in 1950, after which they concluded “that kwashiorkor is the most serious and widespread nutritional disorder known to medical and nutritional science” (Brock and Autret 1952: 72).

In subsequent years, FAO and WHO sponsored other studies in Central America, Brazil, and southern India, and research units in various parts of the world began to concentrate their efforts on determining the etiology, pathogenesis, treatment, and prevention of kwashiorkor.

The similarities and differences between marasmus and kwashiorkor soon became evident and gave rise to intensive debate. D. B. Jelliffe proposed the term “protein calorie (energy) malnutrition” (PEM) to cover the spectrum of syndromes that range from marasmus to kwashiorkor (Jelliffe 1959). This concept was a major contribution in understanding the variations of this group of nutritional diseases.

A history of kwashiorkor appeared in 1954 (Trowell, Davies, and Dean 1954). It discussed an early description of the disease, written by J. P. Correa in 1908 in the Yucatan. Waterlow, in his recent article on the history of kwashiorkor, found an even earlier description from Mexico by F. Hinojosa in 1865 (Waterlow 1991: 236). H. C. Trowell and his colleagues (1954) made it apparent that up to 1954, kwashiorkor had a worldwide distribution. They supported their contention by listing approximately 250 publications

from Africa, Asia, Europe, North and Central America, and South America that contained details on established or probable cases of the disease (Trowell 1954). Most of these dated from between 1920 and 1950, and some early reports from German workers on *Mehlnährschaden* are of especial interest (Czerny and Keller 1925–8). This term, which is best translated as “damaged by cereal flours,” was used to indicate infant malnutrition resulting from imbalanced (excess starch) feeding habits, and the researchers described a clinical picture similar to that of kwashiorkor. In Germany at the time, it had become popular to use cereal gruels instead of milk when a child had gastrointestinal difficulties. The gruels were usually made without milk, and the return of loose stools when milk was once again added to the diet was too often regarded as an indication of sensitivity to milk. The many similarities between *Mehlnährschaden* and kwashiorkor in terms of etiology, pathology, and treatment suggest that these German researchers – in a developed country – may well have been describing kwashiorkor in the early part of the century.

In their book on kwashiorkor, Trowell and colleagues (1954) present a fascinating historical description of the puzzling features of the disease that had confounded investigators and triggered controversies. Among these were the similarities of kwashiorkor to pellagra, the fact that many cases did not have skin rashes, that fatty livers were found in nearly all cases, the role intestinal parasites might play in the disease, and the various forms of treatment. However, what emerged clearly was that lack of protein in the diet was an important cause – a lack caused by dependency on foods that supply adequate carbohydrate but little protein, such as cassava and cereal foods like rice, corn, millet, and sorghum. A distinction between kwashiorkor and infantile pellagra was based on a difference in the distribution of dermatosis (in kwashiorkor the diaper area; in pellagra on exposed areas such as the face, hands, and feet) and a failure of kwashiorkor victims to respond to nicotinic acid unless a high-protein diet was simultaneously given (Trowell et al. 1954: 118–19).

The treatment recommended for kwashiorkor in 1954 was cow’s milk, given in a concentrated form with little lactose and less fat. However, the exact nature of the factor(s) responsible for bringing about improvement in the children’s condition had not been determined. Serum biochemistry showed that treatment increased serum albumin, cholesterol, non-specific esterase, and cholinesterase. But the completeness of recovery and the ultimate prognosis could not at that time be assessed. Moreover, the authors felt that many children suffered from a mild form of kwashiorkor that could not be accurately defined and needed much further investigation (Trowell et al. 1954).

Earlier, the distinction between marasmus and kwashiorkor in tropical countries had been made in

an important monograph, in which children with fatty liver disease (kwashiorkor) were distinguished from those who were undernourished but had no evidence of liver damage (Waterlow 1948). The clinical manifestations of the second group were retarded growth and loss of body fat and resembled the condition known in Europe as infantile atrophy or marasmus. Cases of marasmus were mostly below 60 percent of expected weight for age – the current criterion for a marasmus diagnosis.

Kwashiorkor Since 1954

Following the authoritative reports of Brock and Autret (1952) and Trowell and colleagues (1954), there was intense research into PEM on a worldwide basis. This research was funded by international and national agencies now aware of the high infant and child morbidity and mortality occurring in developing or underdeveloped areas. In addition, academic institutions and individual researchers alike were stimulated to look into such questions. The results have been summarized in recent authoritative publications that have brought our knowledge of PEM up to date (Waterlow 1992; Carpenter 1994).

Dietary Treatment of Kwashiorkor

By 1954, it was clear that milk – as a source of protein – induced recovery, although mortality was still high in seriously ill cases. Brock (personal communication) posed the questions: What was it in milk that brought about recovery? Was it the protein in milk, and if so, what factors or amino acids of that protein initiated it? Were the other constituents of milk, such as the fat, carbohydrate, minerals, trace elements, vitamins, or as yet unknown factors, important for recovery?

A series of clinical trials and balance studies, conducted at Cape Town, South Africa (a nontropical area), from 1953 to 1956, concentrated on what exactly brought about the cure for kwashiorkor and established that a vitamin-free synthetic diet of 11 mostly essential amino acids, glucose, and a mineral mixture could cure the skin rashes, regenerate serum albumin concentration, improve appetite, and eliminate the edema suffered by kwashiorkor victims (Hansen, Howe, and Brock 1956). It was further shown that potassium deficiency as a result of diarrhea and poor intake of potassium-containing foods was an important cause of edema. In fact, edema could resolve without change in serum albumin concentration if potassium depletion was corrected (Hansen 1956).

These studies ended the mystery of what milk contained that initiated recovery from the disease by establishing that protein (amino acid) deficiency was an essential feature of kwashiorkor and that there was no unknown factor involved. Nonetheless, energy deficit was subsequently emphasized by many authors, but often children with kwashiorkor had

enjoyed adequate energy intake and, as with all nutritional deficiency disorders, concurrent vitamin, mineral, and trace-element deficiencies can cause added complications. Milk contains all of these elements except for iron and was thus shown to be an ideal food with which to treat children suffering from kwashiorkor.

Because milk is not universally available, however, there was much interest during the 1950s in the question of whether plant proteins could provide a satisfactory substitute. Nitrogen balance studies during recovery from kwashiorkor (Hansen et al. 1960) revealed that nitrogen (protein) retention was very strong with a milk diet. On a maize (corn) diet, nitrogen retention was much less, but it was greatly improved by the addition of the amino acids missing in maize (lysine and tryptophan), or a legume (pea flour), fish flour, or milk (Hansen 1961). The Institute for Nutrition in Central America and Panama (INCAP) successfully developed a mixture of corn, sorghum, cottonseed flour, and yeast that had good results. It was commercially produced as “Incaparina” in Guatemala and other Central American countries, but it became too expensive for the people who needed it most (Carpenter 1994: 173–5). The extensive study and work with “Incaparina” did, however, prove that commercial vegetable mixtures could be used as a weaning food to promote growth and prevent kwashiorkor.

Between 1955 and 1975 there were numerous other efforts to find substitutes for milk (Carpenter 1994: 161–79). A Protein Advisory Group (PAG) was established by WHO in 1955 and subsequently supported by FAO and the United Nations International Children’s Emergency Fund (UNICEF) to stimulate worldwide research into high-protein foods that might close the so-called protein gap between developed and underdeveloped countries. However, in 1974, a challenging article by D. S. McLaren called into question the importance of protein in the prevention of PEM and stressed that energy depletion was at least as important as lack of protein, if not more so. He argued that marasmus was a more widespread disease than kwashiorkor and that too much emphasis had been placed on – and too much money invested in – the production of protein-rich food mixtures, “. . . whilst children were lost in the unchecked scourge of malnutrition” (McLaren 1974: 95). Although much controversy followed this article (Carpenter 1994: 180–203), the emphasis on the production of high-protein foods waned, and interest became focused on improving food quantity rather than quality (Waterlow and Payne 1975).

Unfortunately, the debate that continues to the present day on the relative importance of protein and energy has often lost sight of the earlier concepts of kwashiorkor and marasmus. Marasmus implied wasting from overall energy lack or starvation, whereas kwashiorkor was characterized by a low-protein diet

that frequently had an adequate energy component. In between the two extremes is marasmic kwashiorkor, which has features of both. Milder cases that manifest only growth retardation (Table IV.D.7.1) can result from a lack of either protein or energy or a combination of the two. A current explanation of the dietary background of PEM is that variations in energy intake, total protein intake, and “quality” of protein (the amino acid pattern) are responsible for the individual clinical forms of PEM.

A protein intake of less than the minimal requirements will result in low serum proteins (hypoalbuminemia) and failure of growth, even in the presence of adequate energy intake (Hansen 1990). Unfortunately, in many parts of the world, most of the dietary protein comes from a single source, often a cereal. Cereals have the disadvantage of being low in total protein content and lacking in essential amino acids, such as lysine (in the case of wheat) or both lysine and tryptophan (in the case of maize). Populations subsisting solely on these foods are thus at risk of energy and protein depletion, and children in particular are at risk of PEM in one form or another. Inevitably, vitamin, mineral, and trace-element deficiencies can complicate PEM in varying degrees, as does infection.

In a review of much new work on protein and energy requirements, Waterlow (1992: 229–59) has concluded that contrary to much that has been published in recent years, some weaning diets in developing countries contain marginal amounts of protein, even when consumed in quantities that satisfy children’s energy needs. Such a marginal diet may satisfy the protein needs of many – perhaps most – children, but not all. Any group of children, as of adults, appears to have a range of protein (and energy) requirements. On marginal intakes, children at the upper end of the range will be at risk. This does not conflict with a controversial finding (Gopalan 1968) that there was no difference, quantitative or qualitative, between the diets of children who developed kwashiorkor or marasmus.

The Liver and Kwashiorkor

A well-described characteristic of kwashiorkor is a fatty infiltration of the liver (Williams 1933). As already mentioned, Waterlow even described kwashiorkor as “fatty liver disease” and distinguished it from cases of undernourishment (marasmus) that showed no evidence of fatty infiltration (Waterlow 1948). In 1945, the Gillman brothers in South Africa published a paper on the successful treatment of fatty liver with a powdered stomach extract, “Ventriculan” (Gillman and Gillman 1945). They referred to their patients as cases of infantile pellagra and, using the liver biopsy technique, observed that such infants had greater or lesser amounts of fat in the liver (without infection) and that the fat accumulation resolved with successful treatment. The Gillmans felt that

“Ventriculan” supplied an essential substance, but other investigators could not confirm their findings and speculated that it was the protein in the diet that was producing the cure.

At that time, there was a high prevalence of cirrhosis of the liver in Africa, and it was thought that suffering PEM early in life might be an underlying cause. This suggestion was refuted, however, by a five-year follow-up study of kwashiorkor cases in a non-tropical environment, which demonstrated that there was complete recovery of the liver with no residual cirrhosis (Suckling and Campbell 1957). In 1969, it was found that there was a connection between fatty liver and serum lipoprotein concentrations (Truswell et al. 1969), and it was hypothesized that fat accumulates in the liver because of the failure of fat transport out of the liver – a failure resulting from the impaired synthesis of apolipoprotein B, which, in turn, is a consequence of protein deficiency. This remains the current view (Waterlow 1992: 61–5).

PEM and Infection

A classic work on the synergistic association of malnutrition and infection was published in 1968 (Scrimshaw, Taylor, and Gordon 1968). In the case of PEM, the condition predisposes to other diseases, but diseases can also bring on the condition. For example, children with PEM are particularly susceptible to respiratory and gastrointestinal infections, whereas measles frequently precipitates severe PEM. An intriguing question is whether PEM interferes specifically with the protective immune responses or whether the generally poor environmental conditions associated with PEM (which implies frequent exposure to infection) means the child with PEM has less metabolic reserve to resist infection. For example, although exposure to measles will infect a well-nourished child and a child with PEM equally, the well-nourished child will survive with some weight loss, whereas the child with PEM, already underweight, becomes severely ill and frequently dies.

The various and complex ways in which immunity to infection can be impaired by PEM has been recently reviewed (Waterlow 1992: 290–324). Although in some communities the relationship between PEM and infection is linear, in others there is a much weaker association. But what is clear is that a child with severe PEM is seriously at risk of infection in any community. An early observer of reduction of cell-mediated immunity in PEM (Smythe et al. 1971) also noted reduction in the weight of the thymus gland as well as a reduced size of the spleen, lymph nodes, tonsils, appendix, and Peyer's patches.

Much interest has also been shown recently in the role of vitamin A deficiency in the susceptibility of PEM victims to infection, especially respiratory disease and diarrhea (Sommer, Katz, and Tarwotjo 1984). Earlier, it was demonstrated that some patients with

kwashiorkor had dangerously low levels of plasma vitamin A (Konno et al. 1968), and treatment of measles with large doses of vitamin A has given good results (Hussey and Klein 1990), which, in conjunction with widespread immunization, means that measles is no longer the threat to the life of PEM victims that it was in 1969 (Morley 1969).

Diarrhea has always been a clinical characteristic of kwashiorkor and marasmus, both as a precipitating factor in a marginally malnourished child and as a continuing recovery-retarding drain of electrolytes, energy, and protein. The organisms and viruses responsible have been well defined (Waterlow 1992: 297), but in the case of PEM victims, frequently no pathogens are isolated. Balance studies on patients recovering from kwashiorkor have revealed a remarkably high daily fecal loss (500 to 1,000 grams [g] per day compared with a normal figure of 100–150 g/day), part of which was found to have been caused by lactose intolerance as a result of secondary lactase deficiency in the duodenum (Bowie, Brinkman, and Hansen 1965; Bowie, Barbezat, and Hansen 1967).

At the time, this discovery was thought to be a breakthrough in the cause of the diarrhea in PEM; further experience, however, revealed that lactose intolerance is not universal in PEM, although it can explain the severe diarrhea that frequently occurs when PEM cases are treated with milk. Diarrhea also occurs in PEM because the gastrointestinal tract atrophies and becomes paper-thin and almost transparent. The mucosa of the intestine has a reduced absorptive surface, and electron microscopy reveals considerable disorganization of the intracellular architecture (Shiner, Redmond, and Hansen 1973). Marked improvement occurs within a few days of treatment.

Looking back on the last 50 years of research on diarrhea in PEM, it is apparent that infection, intestinal atrophy, lactose intolerance, and immunological deficiencies all play their part. Recently, the advent of AIDS has particularly affected the immunological defenses in infected children, resulting in diarrhea, severe wasting, and marasmus or kwashiorkor. In a summing up of all the recent evidence, it can be said that there is a causal relationship between a state of malnutrition (PEM) and diarrhea morbidity and mortality (Waterlow 1992: 313, 339). The same may be said for respiratory disease (pneumonia) and measles, but not for malaria, which has little or no relation to the state of nutrition (Waterlow 1992: 333). Confounding factors in morbidity and mortality are vitamin A deficiency, breast feeding, sanitary facilities, and the mother's education, caring capacity, and availability.

PEM and Body Composition

The profound physical changes in marasmus (wasting) and kwashiorkor (edema) stimulated research into body composition when new techniques became

available after 1950. In Waterlow's (1992) extensive review of the subject, he shows the inconsistencies between different studies and points out that there is still no agreed-upon understanding of the mechanisms of fluid-retention edema. There is a considerable loss of muscle mass and of fat, particularly in marasmus, and as a result, there is an increase of total body water as a percentage of body weight both in kwashiorkor with edema and in marasmus without edema.

Based on evidence available so far, the difference between the two could be that children with edema have more extracellular fluid, which is probably related to the extent of potassium depletion (Mann, Bowie, and Hansen 1972). Kwashiorkor children with edema have lower total body potassium than marasmus cases without edema. It is of interest that the increase of total body water – and of extracellular water as a percentage of body weight – represents a reversion to an earlier stage of development. This means that the weanling child with PEM has the composition and size of a younger child (Hansen, Brinkman, and Bowie 1965). Total body protein is severely depleted in PEM victims, and compared with normal children of the same height, there is a greater deficit of total protein than of body weight. Cellular protein is greatly depleted, but collagen (structural protein) is little affected (Picou, Halliday, and Garrow 1966). The brain is relatively well preserved when compared with other organs in PEM. However, computed tomography has recently shown there is some reversible shrinkage of brain mass (Househam and De Villiers 1987).

PEM in General

Growth Retardation and PEM

A constant feature of kwashiorkor has been growth retardation, occurring in children even before the disease is recognized. Weight, height, and bone development are all affected. In the second half of the twentieth century, anthropometric indices – weight, height, weight-for-height, arm circumference, and skinfold thickness – have been greatly refined and used extensively in the assessment of health and disease. In the Wellcome classification of PEM (Table IV.D.7.1), weight is used as a basis of defining differences between the various syndromes, which has proved most valuable in comparing PEM in different communities and countries.

F Gomez, R. Ramos-Galvan, S. Frenk, and their colleagues in Mexico were the first to divide deficits in weight-for-age into three categories of severity, based on Harvard growth standards (Gomez et al. 1956). This classification had the drawback that it combined in one number the figures for height-for-age and weight-for-height (Waterlow 1992: 189) and was not widely adopted. The Harvard growth charts for height and weight of children in the United States proved

valuable as a standard against which to compare the growth of children with PEM. They were used in the initial Wellcome classification of PEM (Table IV.D.7.1), although later they were superseded by charts from the U.S. National Center for Health Statistics (NCHS), which – though similar – were more thoroughly worked out in terms of statistics. These charts were accepted and published by WHO as international standards (WHO 1983).

There have, however, been controversies concerning the use of these standards, ranging from questions about possible ethnic and environmental influences on growth to debates over the desirability of using national growth charts. Yet environmental and income differences have enormous effects in local surveys (Wittmann et al. 1967), which makes the use of local standards impractical, especially as there is often a secular trend toward improvement in disadvantaged groups. Currently, for the community and individual assessment of children with PEM, cutoff points on the international charts are employed; for example, children who are below the third percentile, or 2 standard deviations below the mean figures for height, weight, or weight-for-height indices, are suspect for PEM (Waterlow 1992: 212–28).

In clinic and field assessments, the “Road to Health” weight charts have revolutionized preschool health assessment even among the most unsophisticated populations. These charts are issued to mothers at the birth of their children and updated at each visit to a primary care center, physician, or hospital. D. Morley (1973) had demonstrated the value of continuously monitoring weight gain and of making the mother responsible for keeping the record. This is an interesting example of the practical procedures that grew out of the early observations of the growth of African children with kwashiorkor (Trowell et al. 1954: 70–3).

The retardation of growth caused by kwashiorkor immediately raised questions about its reversibility. Provided there were improvements in nutrition and environment, could a preschool child regain his or her genetic potential for growth? An early prospective study, started in 1959, monitored children admitted to hospitals with kwashiorkor for a subsequent period of 15 years (Bowie et al. 1980). This study showed that growth retardation resulting from severe PEM is reversible if environment and food intake are reasonably adequate during the prepubertal years. In similar studies, the increment in height was very much the same as that achieved by children in the United States, regardless of the degree of stunting at 5 years (Waterlow 1992: 195–211). As Waterlow has discussed, however, where catch-up has not occurred, stunting caused by long-continued protein and energy lack can lead to functional consequences, for example, a reduction in absolute capacity for physical work. For a given workload, people who are small, even though fit, are at a disadvantage. Thus, research

on body growth in PEM has gone on to stimulate investigation into many interesting aspects of human development and function.

PEM and Intellectual Development

In 1963, a pioneering follow-up study of marasmic infants in South Africa focused on the possible effects of PEM on brain size and intellectual development (Stoch and Smythe 1963). This sparked ongoing worldwide research and intense political interest. Decreased brain weight in a state of malnutrition was reported from East Africa in 1965 (Brown 1965), and, as mentioned earlier, computed tomography has demonstrated that there is a reversible shrinkage of brain mass in kwashiorkor (Househam and De Villiers 1987). The more difficult assessment of the effect of PEM on intellectual development is confused by the interaction of nutrition per se and other environmental factors such as poor social conditions, nurturing, education, and environmental stimulation. In addition to protein and energy, other nutrients – iron, potassium, trace elements, or vitamins, to name a few – may be deficient, and this, too, may affect mental development (Grantham-McGregor 1992).

There is, however, some evidence that PEM does not necessarily cause permanent damage to the intellect. Follow-up studies have failed to demonstrate differences in intellectual development between ex-kwashiorkor patients, their nonaffected siblings, and other children from the same environment (Evans, Moodie, and Hansen 1971). Planned stimulation in hospital has produced recovery of cognitive development (to a normal level) in kwashiorkor children (Cravioto and Arrieta 1979: 899), and a recent study in Jamaica showed that a nutritional supplement provided significant benefits to stunted children between 9 and 24 months of age, as did stimulation alone – but the two together had the best result (Grantham-McGregor 1992).

A WHO symposium in 1974 concluded that in spite of the widely held opinion that PEM in early life permanently jeopardizes mental development, the evidence to support this contention was scanty. Twenty-three years later, this still appears to be the case, but it does seem probable that there is an interaction between malnutrition and other environmental factors, especially social stimulation, and that a child's intellectual status is the result of this interaction. However, good nutrition in the first two years of life enables an underprivileged child to make better use of what stimulus there is in the environment. This kind of beginning has a long-lasting effect on intelligence even if nutrition after this period is less than optimal (Evans et al. 1980).

The Social Background of PEM

Trowell and colleagues (1954: 49–51) pointed out that PEM has always been associated with poor communities and occurs among the most depressed social

classes. These researchers also mentioned social customs at the time of weaning, and of course Williams interpreted the word kwashiorkor to mean the “disease of the displaced child” – displaced by the next baby (Williams 1935). In addition to these basic factors, overpopulation and the movement of rural people to urban areas were considered important. A. Moodie (1982), a social worker who devoted her working life to the study of the background of PEM, as seen in a nontropical area of Africa, has reviewed the literature, including some of her own studies. She noted the following as constituting the essential background of PEM:

1. Economic inadequacy or poverty – rural or urban.
2. Lack of sophistication and knowledge loosely termed ignorance and the cultural factors underlying this state.
3. Problems of overpopulation and too large families. (Here the mother suffers and it is through the mother that child nutrition is mediated. In all programs for prevention the well-being of the mother should be a priority.)
4. Social disorganization, especially illegitimacy and alcohol abuse.
5. High incidence of infection and diarrhea.
6. Early weaning and decline of breast feeding.

In an urban study funded by WHO, low income was found to be critical (Wittmann et al. 1967). In another investigation, sociological and cultural factors such as the working mother appeared to be more important (Shuenyane et al. 1977). There have been many studies throughout the world showing what Waterlow has termed the multiplicity of causes. He grouped them under three headings: lack of food, infection, and psychosocial deprivation (Waterlow 1992: 9–11). It has been pointed out that economists and other planners are now recognizing that nutritional indicators (for example, growth retardation or the clinical features of PEM) provide a more sensitive, objective, and easily collected measure of socioeconomic development than conventional indicators such as per capita income (Church, Burgess, and Burgess 1991: 569–70).

Prevention of PEM

With the high prevalence of PEM in some areas of developing countries (marasmus afflicts 5 to 10 percent and stunting 30 to 60 percent of the under-5 population), prevention has received much attention from international and government agencies. These organizations have been aided by economists and social scientists as well as by nutritionists and health scientists, and the literature has grown enormously (Waterlow 1992: 361–92).

The basic strategy for achieving “health for all” by the year 2000 is Primary Health Care (PHC). This plan includes growth monitoring, health education, maternal and child health care, family planning, immuniza-

tion against major infantile diseases, and appropriate treatment of common diseases, like oral rehydration treatment for diarrhea. There is no doubt that these measures, implemented by PHC teams, have had an overall effect on preventing PEM, which has shown a decline in prevalence in many countries (Bellamy 1996: 54). The details of nutrition inputs in primary care are well described in a recent publication (Church et al. 1991). The problem of PEM needs a *holistic* approach. Health teams and public-health authorities can reduce mortality and morbidity with active programs of clinical care, rehabilitation, and food supplementation, but they cannot affect the prevalence of PEM resulting from underlying socioeconomic and cultural realities. Physicians must have the active, integrated, and effective cooperation and assistance of economists, agriculturalists, and governments to eradicate PEM.

Treatment of PEM

In a remarkable but little-quoted paper, published in an obscure journal, skimmed lactic acid milk was shown to reduce the 40 to 60 percent mortality of severe kwashiorkor to 20 percent (Altmann 1948). The author stressed the importance of small feeds initially until appetite returned. He also stressed the danger of severe dehydration that causes 5 out of 6 of the deaths. A colleague at the same hospital later found that the use of intravenous fluids in dehydrated cases cut mortality by half (Kahn 1959: 161–5). The same principles of feeding and rehydration are followed today (Waterlow 1992: 164–86), with perhaps more emphasis on oral rehydration. Refinements include supplements of potassium, magnesium, zinc, vitamin A, folic acid, and iron. Blood transfusion is used only for very severely anemic children. Because of the frequency of infections, antibiotics are routinely given and, in tropical areas, malaria and other infestations have to be dealt with, using appropriate therapy and prophylaxis. Mortality of severe PEM cases should now be less than 10 percent. Less severe cases of PEM respond well – with negligible mortality – to diets providing adequate energy and protein.

History of the Cause of PEM

Waterlow (1992) has concluded that there is no reason to abandon the concepts put forward many years ago, namely, that kwashiorkor develops when the diet has a low protein-energy (P/E) ratio and that when energy is limiting, the end result is marasmus. To this it should be added that individual children vary in their requirements of nutrients. If the P/E ratio is marginal, protein or energy may be a limiting factor for some children but not for others. Research has shown that total protein intake and the quality of the protein (the amino acid pattern) are also important. In areas where PEM occurs, vitamin, mineral, and trace-element deficiencies can complicate the basic syndrome to varying degrees, as does infection.

Recently, it has been hypothesized that all the serious features of kwashiorkor – edema, fatty liver, infection, and mortality – can be explained by an excess of free radicals (Golden and Ramdath 1987). This theory has not yet been firmly established, and in any case, children still have to be short of protein or energy before they become susceptible to free-radical excess. Another theory is that kwashiorkor results from aflatoxin poisoning (Hendrickse 1984). Aflatoxins come from fungi growing on improperly dried nuts. However, aflatoxins cannot be blamed for kwashiorkor and marasmus that occur where there is no aflatoxin contamination of the diet.

PEM in Perspective

In the nineteenth and twentieth centuries, PEM, as a particularly important nutritional deficiency in infants and children, has become recognized and better understood. This has come about through extensive worldwide observation and research, and recent emphasis has been on the effect of PEM on health, growth, and intellectual development.

As we enter the twenty-first century, we have the knowledge, if not always the means, to limit the prevalence of PEM in individuals and communities at risk. A looming danger is the “demographic trap,” the situation that arises when population growth exceeds growth in food production or availability (King and Elliot 1994: 528). Signs of the trap are already present in several countries, particularly in Africa (Bonneux 1994). If this situation is not addressed by fertility control, PEM is likely to remain with us for the foreseeable future.

J. D. L. Hansen

Bibliography

- Altmann, A. 1948. The syndrome of malignant malnutrition (kwashiorkor, infantile pellagra). Its conception as a protein deficiency and its treatment with skimmed lactic acid milk. *Clinical Proceedings (Cape Town)* 7: 32–53.
- Beaven, P. 1957. A history of the Boston Floating Hospital. *Pediatrics* 19: 629–38.
- Bellamy, C. 1996. *The state of the world's children*. Oxford.
- Bonneux, L. 1994. Rwanda: A case of demographic entrapment. *Lancet* 344: 1689–90.
- Bowie, M. D., G. O. Barbezat, and J. D. L. Hansen. 1967. Carbohydrate absorption in malnourished children. *American Journal of Clinical Nutrition* 20: 89–97.
- Bowie, M. D., G. L. Brinkman, and J. D. L. Hansen. 1965. Acquired disaccharide intolerance in malnutrition. *Journal of Pediatrics* 66: 1083–91.
- Bowie, M. D., A. D. Moodie, M. D. Mann, and J. D. L. Hansen. 1980. A prospective 15-year follow-up study of kwashiorkor patients. Part I. Physical growth and development. *South African Medical Journal* 58: 671–6.
- Brock, J. E., and M. Autret. 1952. *Kwashiorkor in Africa*. Geneva.
- Brown, R. E. 1965. Decreased brain weight in malnutrition and

- its implications. *East African Medical Journal* 42: 584–95.
- Carpenter, K. J. 1994. *Protein and energy: A study of changing ideas in nutrition*. Cambridge.
- Chapin, H. D. 1915. A plea for accurate statistics in infants' institutions. *Archives of Pediatrics* 32: 724–6.
- Church, M. A., A. Burgess, and L. Burgess. 1991. Nutrition in primary care in developed and developing countries. In *Textbook of pediatric nutrition*, Third edition, ed. D. S. McLaren, D. Burman, M. R. Belton, and A. F. Williams, 557–91. Edinburgh.
- Cravioto, J., and R. Arietta. 1979. Stimulation and mental development of malnourished infants. *Lancet* 2: 899.
- Cullen, W. 1789. *A treatise of the materia medica*. 2 vols. Edinburgh.
- Czerny, A., and A. Keller. 1925–8. *Des Kindes Ernährung, Ernährungs Störungen Ernährungs Therapie*. Second edition. Leipzig.
- Darrow, D. C. 1946. The retention of electrolytes during recovery from severe dehydration due to diarrhoea. *Journal of Pediatrics* 28: 515–40.
- Dickens, C. [1858] 1956. Speech on behalf of the Hospital for Sick Children. *Journal of Pediatrics* 49: 607.
- Evans, D., M. D. Bowie, J. D. L. Hansen, et al. 1980. Intellectual development and nutrition. *Journal of Pediatrics* 97: 358–63.
- Evans, D., A. D. Moodie, and J. D. L. Hansen. 1971. Kwashiorkor and intellectual development. *South African Medical Journal* 45: 1413–26.
- Gillman, T., and J. Gillman. 1945. Powdered stomach in the treatment of fatty liver and other manifestations of infantile pellagra. *Archives of Internal Medicine* 76: 63–75.
- Golden, M. H. M., and D. Ramdath. 1987. Free radicals in the pathogenesis of kwashiorkor. *Proceedings of the Nutrition Society* 46: 53–68.
- Gomez, E., R. Ramos-Galvan, S. Frenk, et al. 1956. Mortality in second and third degree malnutrition. *Journal of Tropical Pediatrics* 2: 77–83.
- Gopalan, C. 1968. Kwashiorkor and marasmus: Evolution and distinguishing features. In *Calorie deficiencies and protein deficiencies*, ed. R. W. McCance and E. M. Widdowson, 49–58. London.
- Govan, C. D., and D. C. Darrow. 1946. The use of potassium chloride in the treatment of the dehydration of diarrhoea in infants. *Journal of Pediatrics* 28: 541–9.
- Grantham-McGregor, S. M. 1992. The effect of malnutrition on mental development. In *Protein energy malnutrition*, ed. J. C. Waterlow, 344–60. London.
- Habicht, J. P., R. Martorell, C. Yarbrough, et al. 1974. Height and weight standards for pre-school children. How relevant are ethnic differences in growth potential? *Lancet* 1: 611–15.
- Hansen, J. D. L. 1956. Electrolyte and nitrogen metabolism in kwashiorkor. *South African Journal of Laboratory and Clinical Medicine* 2: 206–31.
1961. The effects of various forms of supplementation on the nutritive value of maize for children. In *Meeting protein needs of infants and children*. National Academy of Sciences, National Research Council, Pub. No. 843, 89–102. Washington, D.C.
1990. Malnutrition review. *Pediatric Reviews and Communications* 4: 201–12.
1991. Marasmus. In *History of pediatrics, 1850–1950*, ed B. L. Nichols, A. Ballabriga, and N. Kretchmer, 219–31. New York.
- Hansen, J. D. L., G. L. Brinkman, and M. D. Bowie. 1965. Body composition in protein calorie malnutrition. *South African Medical Journal* 39: 491–5.
- Hansen, J. D. L., E. Howe, and J. F. Brock. 1956. Amino acids and kwashiorkor. *Lancet* 2: 911–13.
- Hansen, J. D. L., and J. M. Pettifor. 1991. Protein energy malnutrition (PEM). In *Textbook of pediatric nutrition*. Third edition, ed. D. S. McLaren, D. Burman, N. R. Belton, and A. F. Williams, 357–90. Edinburgh.
- Hansen, J. D. L., H. E. Schendel, J. A. Wilkins, and J. F. Brock. 1960. Nitrogen metabolism in children with kwashiorkor receiving milk and vegetable diets. *Pediatrics* 25: 258–82.
- Hendrickse, R. G. 1984. The influence of aflatoxins on child health in the tropics with particular reference to kwashiorkor. *Transactions of the Royal Society of Tropical Medicine and Hygiene* 78: 427–35.
- Holt, L. E., A. M. Courtney, and H. E. Fales. 1915. The chemical composition of diarrhea as compared with normal stools in infants. *American Journal of Diseases of Childhood* 9: 213–24.
- Househam, K. C., and F. J. K. De Villiers. 1987. Computed tomography in severe PEM. *Archives of Disease in Childhood* 62: 589–642.
- Hussey, G. D., and M. Klein. 1990. A randomized controlled trial of vitamin A in children with severe measles. *New England Journal of Medicine* 323: 160–4.
- Jelliffe, D. B. 1959. Protein-calorie malnutrition in tropical pre-school children: A review of recent knowledge. *Journal of Pediatrics* 54: 227–56.
- Kahn, E. 1959. Prognostic criteria of severe protein malnutrition. *American Journal of Clinical Nutrition* 7: 161–5.
- Kerpel-Fronius, E. 1947. Infantile mortality in Budapest in the year 1945. *Journal of Pediatrics* 3: 244–9.
1983. *The pathophysiology of infantile malnutrition*. Budapest.
- King, M., and C. Elliot. 1994. Cairo: Damp squibs or Roman candle? *Lancet* 344: 528.
- Konno, T., J. D. L. Hansen, A. S. Truswell, et al. 1968. Vitamin A deficiency and protein calorie malnutrition in Cape Town. *South African Medical Journal* 48: 950–5.
- Mann, M. D., M. D. Bowie, and J. D. L. Hansen. 1972. Potassium in protein calorie malnutrition. *South African Medical Journal* 46: 2062–4.
- Martorell, R. 1985. Child growth retardation: A discussion of its causes and its relationship to health. In *Nutrition adaption in man*, ed. K. Blaxter and J. Waterlow, 13–30. London.
- McLaren, D. S. 1974. The protein fiasco. *Lancet* 2: 93–6.
- Moodie, A. 1982. Social background of child nutrition. In *Textbook of pediatric nutrition*. Second edition, ed. D. S. McLaren and D. Burman, 349–63. New York.
- Morley, D. 1969. Severe measles in the tropics. *British Medical Journal* 1: 297–300, 363–5.
1973. *Pediatric priorities in the developing world*. London.
- Picou, D., D. Halliday, and J. S. Garrow. 1966. Total body protein, collagen and non-collagen protein in infantile protein malnutrition. *Clinical Science* 30: 346–51.
- Scrimshaw, N. S., C. E. Taylor, and J. E. Gordon. 1968. *Interactions of nutrition and infection*. WHO Monograph Series, No. 57. Geneva.
- Shiner, M., A. O. B. Redmond, and J. D. L. Hansen. 1973. The jejunal mucosa in protein energy malnutrition. A clinical, histological and ultrastructural study. *Experimental and Molecular Pathology* 19: 61–78.
- Shuenyane, E., S. Mashigo, C. Eyberg, et al. 1977. A socio-economic, health and cultural survey in Soweto. *South African Medical Journal* 51: 495–500.

- Smith, C. A. 1983. *The Children's Hospital of Boston*. Boston, Mass.
- Smythe, P. M., M. Schonland, G. G. Brereton-Stiles, et al. 1971. Thymolymphatic deficiency and depression of cell-mediated immunity in protein-calorie malnutrition. *Lancet* 2: 939–44.
- Sommer, A., J. Katz, and I. Tarwotjo. 1984. Increased risk of respiratory disease and diarrhea in children with preexisting mild vitamin A deficiency. *American Journal of Clinical Nutrition* 40: 1090–5.
- Stoch, M. B., and P. M. Smyth. 1963. Does undernutrition during infancy inhibit brain growth and subsequent intellectual development? *Archives of Disease in Childhood* 38: 546–52.
- Suckling, P. V., and J. A. H. Campbell. 1957. A five-year follow-up of coloured children with kwashiorkor in Cape Town. *Journal of Tropical Pediatrics* 2: 173–80.
- Trowell, H. C., J. M. P. Davies, and R. F. A. Dean. 1954. *Kwashiorkor*. London.
- Truswell, A. S., J. D. L. Hansen, C. E. Watson, et al. 1969. Relation of serum lipids and lipoproteins to fatty liver in kwashiorkor. *American Journal of Clinical Nutrition* 22: 568–76.
- Waterlow, J. C. 1948. *Fatty liver disease in the British West Indies*. Medical Research Council Special Report Series, No. 263. London.
1991. Kwashiorkor. In *History of pediatrics, 1850–1950*, ed. B. L. Nichols, A. Ballabriga, and N. Kretchmer, 233–46. New York.
1992. *Protein energy malnutrition*. London.
- Waterlow, J. C., and P. R. Payne. 1975. The protein gap. *Nature* 258: 113–17.
- Wellcome Trust Working Party. 1970. Classification of infantile malnutrition. *Lancet* 2: 302–3.
- Williams, C. D. 1933. A nutritional disease of childhood associated with a maize diet. *Archives of Disease in Childhood* 8: 423–33.
1935. Kwashiorkor: A nutritional disease of children associated with a maize diet. *Lancet* 2: 1151–2.
- [1963] 1973. The story of kwashiorkor. *Nutrition Reviews* 31: 334–40.
- Wittmann, W., A. D. Moodie, S. Fellingham, and J. D. L. Hansen. 1967. An evaluation of the relationship between nutritional status and infection by means of a field study. *South African Medical Journal* 41: 664–82.
- World Health Organization (WHO). 1974. Malnutrition and mental development. *WHO Chronicle* 28: 95–102.
1983. *Measuring change in nutritional status. Guidelines for assessing the nutritional impact of supplementary feeding programmes for vulnerable groups*. Geneva.

IV.D.8 🐉 Scurvy

The human body requires an adequate supply of ascorbic acid (L-xyloascorbic acid or vitamin C) to enable it to function normally, and a lack of the vitamin results in the emergence of the condition known as scurvy (scorbutus or avitaminosis C). Unlike plants, and the majority of animals thus far studied, humans are unable to produce ascorbic acid endogenously and, thus, are dependent upon dietary sources – mainly fruit and vegetables – for a supply of the vita-

min. In the absence of vitamin C, formation of collagen, an essential glycoprotein component of connective tissue, is impaired, which is believed to be the main underlying biochemical lesion in scurvy (Counsell and Hornig 1981; England and Seifter 1986).

The earliest signs of scurvy (fatigue and lassitude) may emerge in humans some 12 weeks after removal of dietary vitamin C, and the more overt traditional signs (hemorrhagic spots under the skin [*petechiae*], softening of the gums, and defective wound healing) after some 17 to 26 weeks of deprivation.

In 1753, James Lind concluded his pioneer study of scurvy with a chronological *Bibliotheca Scorbutica*, which imparted a mild historical flavor to his text (Stewart and Guthrie 1953). But more than a century was to pass before the first sustained efforts to produce a history of the disease emerged. One of these was J. Maré's 200-page article in the *Dictionnaire Encyclopédique des Sciences Médicales* in 1880, and the second was August Hirsch's 60-page article in his *Handbook of Geographical and Historical Pathology* (1883–6). Both of these pioneer works are now completely eclipsed by K. J. Carpenter's comprehensive *History of Scurvy and Vitamin C* (1986), to which the reader is referred for a more extended treatment.

Simple logic dictates that scurvy became a clinical possibility when evolving hominoids (unlike most other animals) lost the ability to biosynthesize vitamin C from carbohydrate precursors. This loss would, presumably, only have occurred in an environment that provided early humans with an adequate intake of dietary vitamin C. By the same token, scurvy itself would not have emerged until a change in food consumption patterns or in the seasonal availability of the vitamin resulted in a deprivation of vitamin C.

The main dietary sources of vitamin C are fruit and vegetables, particularly when eaten raw, and to a much lesser extent, fresh meat (salting or drying meat results in a loss of its vitamin C content). Grains, nuts, eggs, and dairy products (with the exception of fresh milk, which contains small amounts) provide little or no vitamin C.

There can be little doubt that the unstable and fluctuating environmental conditions of primitive humans must have, from time to time, produced highly scorbutogenic situations, but the precise frequency of the disease in prehistoric and early historic times is obscure. Some commentators have claimed to discern descriptions of scurvy in the writings of early classical writers such as Pliny and Paulus Aeginita, but this must remain a matter of considerable conjecture.

There is some evidence that scurvy afflicted the armies of the Crusaders in the thirteenth century, and the writings of Jacques de Vitry (c. 1220) and Jean Sire de Joinville (1224–1319) have been quoted in this respect (Major 1978: 585–6). Indeed, de Joinville's description of the "army sickness" reveals a situation at least as fearful as the depredations of the attacking Turks:

... our legs shrivelled up and became covered with black spots, and spots of the colour of earth, like an old boot: and in such of us as fell sick the gums became putrid with sores, and nor man recovered of that sickness, but all had to die. . . . The proud flesh in our men's mouth grew to such excess that the barbour-surgeons were obliged to cut it off, to give them a chance of chewing their food or swallowing anything. It was piteous to hear through the camp the shrieks of the people who were being operated for proud flesh, for they shrieked like women in childbirth. (Major 1978: 586)

Scurvy as an Occupational Disease

Curiously enough, the change that most clearly and finally precipitated the emergence of scurvy as a recognizable disease was not so much dietary as technological. Scurvy first established itself as a discrete and clearly definable feature of early medicine because of those technological advances that enabled the building of ships capable of prolonged sea voyages.

By the end of the fifteenth century the ship had acquired the structure that was to characterize it for the next three centuries, and the newer naval technology meant that humans could now remain at sea for months at a time – long enough to deprive themselves of an adequate supply of the fresh fruit and vegetables that are the main dietary source of vitamin C. Consequently, it was during the early voyages of exploration in the fifteenth century that the first clear pictures of scurvy emerged. In this sense scurvy (or scorbutus as it was described in the Latin texts of the day) may truly be regarded as one of the earliest occupational diseases.

Perhaps the most widely quoted example of explorer's scurvy is that described by Jacques Cartier, the Breton explorer, who during his discovery of the Saint Lawrence River in 1536 lost 25 of his crew to a "strange and cruell disease." Certainly his description of their condition – the loss of strength and "their skins spotted with spots of blood of a purple colour . . . [and] their gummies so rotten, that all the flesh did fall off, even to the rootes of the teeth, which did also almost all fall out" is consistent with the petechiae and connective tissue defects now accepted as cardinal features of scurvy (Major 1978: 586).

But a similar condition prevailed among the crew of Vasco da Gama during his return voyage from Africa in 1498; and almost a hundred years later Richard Hawkins could assert that during his 20 years at sea he had witnessed some 10,000 cases of scurvy (Carpenter 1986: 1–26).

A century and a half after that, George Anson lost three-fourths of his men and officers, mostly to scurvy, in his 1740–4 voyage around the world. In fact, it has been estimated that during the critical period between 1600 (when long sea voyages had

become more common) and 1800 (when the prophylactic effect of lemon juice rations had proved effective), scurvy was responsible for over 1 million deaths among seamen – almost certainly more than the sum total lost during the same period in naval battles, shipwrecks, and to diseases other than scurvy (Roddiss 1951: 48, 72).

In the seventeenth century scurvy was considered primarily a scourge of sailors and soldiers. "In Navies and Camps, Scorbutick Feavers are very frequent, and arrive to a great height of malignity, sweeping them away like the Plague sometimes" wrote Everard Maynwaringe in 1664, adding "And I remember, in a hard and long voyage at sea, most of our people were very Scorbutick . . . who at our first setting out were sound and healthy" (Maynwaringe 1679: 57).

Much was written about the possible causes of scurvy among sailors. Bad (i.e., cold and wet) air, poor diet, idleness, and a melancholy outlook were viewed as major predisposing causes. Overindulgence in salted meat was held to be a main dietary cause – particularly as this could explain the peculiar prevalence of the condition among sailors. And some commentators introduced their own specific dietary causes. Thomas Willis, for example, blamed sugar, whereas Maynwaringe accused tobacco, although neither explanation gained general acceptance.

"Land Scurvy"

George Budd, one of the more competent writers on scurvy in the early nineteenth century, wrote, "Scurvy has, unquestionably, existed in the north of Europe from the most remote antiquity. That we have no mention of it in the early history of the northern nations must be imputed to the extreme ignorance of the people, especially as regards medicine" (Budd 1840: 58).

It is the case that medieval medical texts and herbals contain no clearly identifiable references to scurvy, although the numerous remedies for loose or "wagging" teeth have sometimes been thought to indicate unrecognized scurvy (e.g., Henslow 1899: 112). Nor does scurvy appear to have been included among the main diseases responsible for hospitalization in medieval England (Clay 1909). Thus, as already suggested, the general consensus would appear to relate the recognition and acceptance of scurvy as a specific clinical lesion to its occurrence among sailors during the fifteenth and sixteenth centuries.

Certainly it was in the sixteenth century that sporadic references to scurvy began to appear in the medical texts and herbals. Thus, although there are no recognizable references to scurvy in the herbals of Leonhard Fuchs (1542), Hieronymus Bock (1552), and Pietro Andrea Mattioli (1565), William Turner's *Herbal* (1568) comments on its appearance on the Continent. And Henry Lyte's translation of R. Dodoens' *Cruydeboeck* (1578) refers to:

Spooneworte [*Cochlearia* sp., scurvy grass] . . . a singular remedie against the disease of the mouth . . . called by the Hollanders and Friselanders Scuerbuyck, against whiche evill it hath bene lately proved to be very good and is in great estimation and much used of the Hollanders and Friseans. (Dodoens 1578: 117–18)

Similarly, J. C. von Kraftheim's *Consiliorum et Epistolarum Medicinalium* (1591: 285) gave dietary precepts to be followed in scurvy. But surprisingly, Philip Barrough's comprehensive *The Method of Physick* (third edition, 1596) contained no reference to scurvy, nor did Felix Plater's *Observationum in Homines Affectibus Plerisque, Corpori & Animo* . . . (1651). However, William Clowes, in his *A Profitable and Necessary Booke of Observations* . . . (1596) described treating two sailors who were "sicke at the sea of the Scorby."

One of the earliest descriptions of scurvy in English can be found in Thomas Johnson's version of J. Gerarde's *Herbal* (1633) as part of the entry for scurvy grass (*Cochlearia officinalis*), the most favored of the antiscorbutic herbs (see following). It related scurvy to its supposed classical "precursors" and is worth quoting in some detail:

The juice of Spoonewort given to drinke in Ale or Beere . . . perfectly cureth the disease called of *Hippocrates*, *Voluulus Hematites*: of *Pliny*, *Stomacace*; of *Marcellus*, *Oscedo* and of the later writers, *Scorbutum* of the Hollanders and Frisians, *Scuerbuyck*: in English, the Scurvie . . . this filthy, lothsome, heavy, and dull disease, which is very troublesome and of long continuance. The gums are loosed, swolne, and exulcerate, the mouth greevously stinking; the thighes and legs are withall very often full of blew spots, not much unlike those that come of bruises; the face and the rest of the body is oftentimes of a pale colour: and the feet are swolne, as in a dropsie . . . a disease haunting the campes, which vexes them that are besieged and pinned up and it seemeth to come by eating of salt meates, which is increased and cherished with the cold vapors of the stone walls . . . *Hippocrates* has written: their gums (saith he) are infected, and their mouthes stinke that have great spleenes or milts and . . . can hardly be cured of this malladie, especially of the ulcers in the legs and blacke spots. The same is affirmed by *Paulus Aegineta* in this third booke, 49 chapter, where you may easily see the difference between this disease and the black jaunders, which many times are so confounded together that the difference or distinction is hard to be known but by the expert chirurgions who oftentimes serving in the ships, as wel her Majesties as merchants, are greatly pestered with the curing thereof. . . . (Gerarde 1633: 402)

By the early seventeenth century references to cases of scurvy among land dwellers were beginning to appear; initially, it was believed that there were two forms of the disease, and references were made to "sea scurvy" and to "land scurvy." This dichotomy continued until it was finally dispelled by Lind in 1753. Such an apparent sudden increase in the incidence of scurvy among land dwellers in Britain during the seventeenth century was commented on by a number of writers. Marchmont Nedham in his *Medela Medicinae* (1665) supported his claim that scurvy was on the increase in England by referring to the Bills of Mortality, which recorded an increase in the deaths from scurvy from 5 in 1630 to over 100 in 1656. Similarly, Gideon Harvey commented in 1675:

Many years it [scurvy] remained on that Coast [i.e., Holland] before we were sensible of it here in England: for there are many Physicians yet living who in the former part of their Practice had so much as heard of the Name of this Disease, whereas within the last 20 or 30 years past it's grown very familiar among us. (Harvey 1675: 211)

There is confirmatory evidence that in Wales, too, scurvy was regarded as a new disease at the beginning of the seventeenth century. Sir Thomas Wiliems of Trefriw, a priest and practicing physician, compiled the manuscript for his projected *Thesaurus Linguae Latinae et Cambrobritannicae* during the period 1604–7; under the entry for the plant "Britanica" he refers to scurvy as "a new disease in our land, Wales" ("clevyt newydd yn ein gwlad ni Cymru") (Hughes 1990).

William Vaughan (1575–1641), also writing from Wales, reflected the growing importance of scurvy in seventeenth-century medicine in his *The Newlanders Cure* (1630). He recommended the tops and leaves of turnips as antiscorbutics; more significant, perhaps, he used contemporary thought on scurvy to illustrate his anti-papist sentiments in his verses "Description of the Catholicke Scurvy ingendered by the Mystery of Iniquity . . .":

..As doth the former scurvy beate
For want of Sunne and Motions heate
Upon the Spleene, the Breathe and Skinne
So doth that Old and Scurvy sinne
Withe Purple Spots go on to Stayne
Both Soule and Body, all for Gaine . . .
(Vaughan 1630: 112)

Whether these writings reflected a true increase in the incidence of scurvy among land dwellers, or merely indicated a greater awareness of its presence, is not known. According to John Floyer (better known for his introduction of a minute watch to measure pulse rate), "Scurvy was a new Name for an old Disease" – and there were others who regarded

scurvy merely as an older disease that had changed its nature (Floyer 1722: 3; King 1970: 149).

Diet and Scurvy

A true increase in the incidence of scurvy would have required a fairly substantial change in the dietary pattern (or in the dietary availability of vitamin C), and there is little evidence that any such substantial and sudden change occurred in Britain at the time. Rather, such knowledge as is available would appear to suggest that the consumption of turnips, cabbage, and other vegetables by the "husbandman" sector of the population of Britain in the seventeenth century would have provided them with adequate amounts of the vitamin (Powicke 1926).

Yet with diets of the poorer classes of the day, one is less certain. And of earlier centuries little is known, although if one accepts the general view that the pre-sixteenth-century working-class diet was predominantly grain-based with supplements of salted and/or dried meat or fish but with very little fresh fruit and vegetables, then scurvy must have been of frequent occurrence (Prentice 1939: 118–36).

Of the fifteenth-century peasant diet, it has been written:

... it seems likely that the winter diet of salt bacon, bread and peas gave little protection against scurvy, so that by the end of the winter most of the poor country people must have been in at least a pre-scorbutic condition. (Drummond and Wilbraham 1964: 77)

However, the comparative absence of any clearly recognizable references to scorbutic afflictions among a population that during the winter period presumably subsisted on a scorbutigenic diet is somewhat puzzling.

It was not until the enthusiasm of the Dutch market gardeners spilled over into seventeenth-century Britain that a general, albeit small, reduction in the scorbutigenic potential of the British diet occurred. Yet this, for some unaccountable reason, would seem to be the period that witnessed an apparent increase in the incidence of scurvy. However, a much more significant change in the British diet was yet to occur. The end of the seventeenth century witnessed the introduction of the potato on a large scale – a crop that was subsequently to become, for a very substantial proportion of the population, the main dietary source of vitamin C and, hence, the prime protector against widespread scurvy.

Potatoes and Scurvy

In 1662 a Committee of the Royal Society considered a proposition "to plant Potatoes through all ye parts of England," and in 1664 Forster pressed the case for widespread cultivation of the potato in his *England's Happiness Increased*. It is not known to what extent Forster's advice was acted upon, although Thomas

Moffet claimed in 1665 that "pottato-roots are now so common and known amongst us, that even the Husbandman buys them to please his Wife" (Moffet 1746: 324). In 1691 Richard Baxter, the religious writer, in his last literary composition (*The Poor Husbandman's Advocate*) suggested that all smallholders should plant a quarter of an acre of potatoes to provide themselves with "a half year's wholesome food" (Powicke 1926).

Nevertheless, it is unlikely that these statements reflected any widespread use of the potato, and it is probable that until the end of the seventeenth century it remained primarily a horticultural curiosity cultivated by that sector of the populace least likely to suffer from dietary inadequacies, a situation reflected by the somewhat esoteric recipes for its use in the cookery books of the period (e.g., Salmon 1696: 263).

In the eighteenth century, however, potatoes were much more widely grown. According to *The Complete Farmer* (1777):

the culture of this plant has, within these last thirty years, been extended to almost every part of England. . . . It is esteemed, and now very generally used at the tables of persons of all ranks; and inestimable for the poor being a cheap and very wholesome food . . . in Ireland, particularly in the province of Munster, they are the principal, and almost the only food of the poor there for almost eight months in the year. (Society of Gentlemen 1777)

The importance of the potato in the social etiology of scurvy has not always been fully appreciated, although Carpenter has clearly underlined the significant role that it played in the history of scurvy, particularly in Britain and in Ireland. Its significance stems from two facts. First, although comparatively low in vitamin C (10 to 20 milligrams [mg] per 100 grams [g]), potatoes were eaten in such large amounts that they frequently accounted for a very high proportion of the vitamin, and for many they represented its sole source. In Ireland, during the nineteenth century, daily consumptions of up to 4.5 to 6.5 kilograms (kg) per person were recorded (Letheby 1870: 26; Salaman 1970: 331), which would have provided an estimated daily intake of about 400 to 600 milligrams of vitamin C – some 40 to 60 times the quantity required to prevent scurvy. In the 1840s, scurvy was, therefore, a natural consequence of the failure of the potato crop. As suggested earlier, unlike potatoes, cereal grains contain no vitamin C, and thus the maize imported to alleviate the famine had no effect at all against scurvy.

It is also important to note that potatoes can be stored for many months with very little loss of fluid (and consequently with a retention of ascorbic acid), thereby providing a source of the vitamin well into the winter when fresh sources are absent. Even today a substantial sector of the population of Britain

obtains the bulk of its vitamin C from potatoes, particularly in the winter months.

By the same token, many of the outbreaks of scurvy in the nineteenth century (as in the Irish Famine of 1845–8 and in the Exeter hospital epidemic in 1846) were directly traceable to a reduced supply of potatoes (Carpenter 1986: 101; Shapter 1847: 945–8). In charting the history of scurvy, a role of paramount significance must be accorded to the potato; its establishment in Italy and Spain some centuries before it attained popularity in Britain was possibly a factor in determining the higher frequency of scurvy in northern Europe in the sixteenth and seventeenth centuries (Salaman 1970: 142–6). During the last three centuries the potato has doubtless protected millions of people from the ravages of the disease.

Early Writers and Theories

Many of the reports of scurvy in Europe in the sixteenth century emphasized its essential prevalence in the colder northeastern areas. “The Scurvy is properly said to be endemical in most of our Northern Countreys, that border upon the Baltick Sea, or adjacent to the German Ocean: As Denmark, Swedland, Norway, Ffrisland, Holland, England etc. But in High Germany, France, Spain and Italy, the Scurvy is accounted sporadical . . . but here and there one” (Maynwaringe 1679: 16). Such was the comparative rarity of scurvy in southern France that Lazarus Riverius, writing “from my study in Montpelier on the 8th of July 1653,” dismissed the condition, commenting, “The scurvy is usual in the North [of France?] but almost unknown in the South . . . (Riverius 1655: 357–61). The non sequitur that a cold and damp atmosphere conduced to scurvy prevailed in Europe until well into the eighteenth century.

A. S. Hess, however, in commenting on the apparent absence of references to scurvy in the classical medical writings, offered a truer explanation: “Greek, Roman and Arabic writers do not seem to have been acquainted with scurvy. This is as we should expect, for fruits and vegetables grew in such plenty in these southern countries that scurvy must have been a disorder of rare occurrence” (Hess 1920: 1). This (together with the comparatively early adoption of the potato in southwestern Europe) is presumably also the explanation of the differential distribution pattern of scurvy reported for Europe in the sixteenth and seventeenth centuries.

A parallel development to this apparent increase in scurvy (or the recognition of it) among land dwellers in parts of northern Europe in the seventeenth century was the appearance of the first medical texts devoted entirely to the condition. Unfortunately, however, it is impossible to determine whether the appearance of the texts resulted in an increased awareness of the condition, and hence an apparent

increase in its incidence, or whether the texts were produced in response to a genuine increase in the frequency of the disease.

In any event, these medical texts first appeared on the Continent, and J. Echthius's *De Scorbuto* (1541) is usually regarded as the pioneer study; the ensuing two centuries saw a substantial proliferation of “scorbatic” texts – Lind in 1753 referred to some 40 of them in his *Treatise*; but few of these early compilations showed any great originality of thought or interpretation. Probably the two most influential were John Wier's *De Scorbuto* published in Basel as a part of his *Medicarum Observationum* . . . in 1567 and Severinus Esgalenus's *De Morbo Scorbuto Liber* . . . first published in 1604 and then again in much changed forms under the hands of subsequent editors up to 1658.

John Wier [Wierus] – better known perhaps for his *De Praestigiis Daemonum* (1563) (a treatise opposing the persecution of witches) – was probably the more scientific of the two. His treatise on scurvy was reprinted in its entirety by Daniel Sennert in 1624 as one of the six component works in his *De Scorbuto* (1624). In the prefatory remarks to an English translation that appeared in 1622, the translator indicated that he had translated Wierus rather than Ronseus or Langius or Echthius because of a tendency on their part to “wade into deep difficulties” and because they “paynteth out the signes, and poynteth to the cure, but affordeth not the pith and marrow of speciall medicines . . .” (Wier 1622: ii).

Wier's *De Scorbuto* was also the basis of the *Traetice of the Scorbute* written in 1676 by Gwilym Pue [Puw], a Welsh Catholic recusant priest and physician. Pue's treatise, incidentally, contained a case history of a supposedly scorbutic patient – probably Pue himself (Hughes In press). S. Esgalenus exerted a similar influence during a slightly later period, and he was freely quoted until well into the eighteenth century, but his work, even in the later “improved” editions, was less well presented and much more anecdotal than that of Wier. Lind was highly critical of Esgalenus and of Willis, the English physician who uncritically represented Esgalenus in his *Tractatus de Scorbuto* (1667).

Despite the considerable attention that it attracted, very little, in reality, was known about scurvy in the seventeenth and eighteenth centuries. There was fairly general agreement about its main clinical features but little was known of its cure and still less of its cause. The traditional belief that it was essentially a disease of the spleen persisted until well into the eighteenth century. The situation was further complicated by a tendency to assume that there were different types of scurvy. As indicated earlier, it was believed that “sea scurvy” was a different condition from “land scurvy”; there were references too to “hot scurvy” and to “cold scurvy,” to “acid scurvy” and to “alkaline scurvy,” and the situation was further con-

fused by the belief that these categories were not mutually exclusive.

Furthermore, there were probably different “grades” of scurvy, corresponding to different degrees of vitamin C deprivation. Scurvy, when it occurred among land dwellers, was presumably often a case of chronic partial vitamin C deficiency (hypovitaminosis C) rather than the “full” scurvy (avitaminosis C) that afflicted sailors after their much longer periods of complete deprivation of the vitamin. Hence the confusion between the clear-cut symptoms of complete deficiency among sailors and the chronic (and often seasonal) hypovitaminosis C that could befall land dwellers.

The “Antiscorbutic” Plants

The classical interpretation that diseases and ill health reflected a change in the humoral balance of the body was a powerful element in seventeenth- and eighteenth-century medical thought. Dietary and pharmacological treatments advocated for diseases were frequently devised primarily to counteract or correct the supposed deviant humoral patterns. Edward Strother, in introducing his influential *An Essay on Sickness and Health* in 1725, was quite categorical on this point: “It is therefore a standing Rule that our Meats and Drinks ought to consist, as nearly as can be, of Particles contrary to the Cause of the Disease reigning . . .” (1725: 30).

Correspondingly, eighteenth-century texts on diet and nutrition attached considerable importance to a categorization of foods in terms of their supposed humoral properties so that physicians could select appropriate foods to counteract specific diseases. “The best way of curing the Gout,” wrote W. Stephens (1732: 60) “is to hinder the Generation of this gouty Humour in the Body; this is to be effected no other way, that I know of, but by Diet.”

The same was to hold true in the case of scurvy. For “hot,” “alkaline,” “sea,” scurvy, cooling acidic foods, such as oranges and lemons, were advocated; “cold,” “acid,” “land,” scurvy, in contrast, could be treated by the “hot” antiscorbutic plants such as scurvy grass, brooklime, and the cresses. Thus there developed a tradition, reinforced by empirical observations, that “sea-scurvy” could be treated by oranges and lemons, whereas the methods of choice for scurvy among land dwellers centered on the antiscorbutic herbs. When John Woodall extolled the virtues of lemon juice as a cure for scurvy among sailors he carefully contrasted its value with that of the traditional antiscorbutics “namely Scurvy-grasse, Horse reddish roots, Nasturtia Aquatica . . . and many other good meanes” whose virtues extended, however, “only to the Cure of those which live at home . . .” (Woodall 1639: 61).

Turner, in describing the virtues of brooklime in his *Herbal* of 1568, penned what must be one of the earliest references to scurvy in an English herbal:

“ . . . I have proved it my selfe by experience that brooklyme is very good for a decease that reigneth much in Freseland called the Scourbuch. I sod the herbe in butter milke, the cheese and butter taken away, and gave the patientes it so” (128). Clearly, this statement lends further support to the belief that scurvy was either comparatively rare or unrecognized in sixteenth-century Britain.

As it happens, the concentration of foliar ascorbic acid in the three antiscorbutic herbs (*Cochlearia officinalis* [scurvy grass], *Veronica beccabunga* [brooklime], and *Nasturtium officinale* [watercress]) – and particularly in scurvy grass – is low when compared with other angiosperms. Although fresh preparations of the “antiscorbutics” would, in sufficient amounts, certainly have cured scurvy, it seems likely that their entry into, and retention by, sixteenth-century medical literature probably reflected a priori thoughts on the humoral nature of “land scurvy” more than any observed genuine therapeutic value.

James Cook provided his crew with a wide range of vegetables and fruit and scurvy grass collected whenever available, and this, no doubt, protected them from the disease. Had Cook, however, relied on fresh scurvy grass alone, a weekly supply of some 100 pounds of scurvy grass leaf would have been required for a ship with a complement of 100, to provide the Dietary Reference Value of vitamin C (40 mg daily) for the prevention of scurvy.

Nevertheless, the antiscorbutic triumvirate retained a position of importance in the herbals of the sixteenth, seventeenth, and eighteenth centuries, although, as has recently been shown, almost all of the original ascorbic acid would have been lost because of the form in which the preparations were ultimately administered – a finding that must cast considerable doubt on their overall effectiveness (Hughes 1990).

A Nosological Safety Net

The picture was further confused in the eighteenth century by a readiness of physicians to describe a whole range of unrelated conditions as “scurvy”; and the term became a convenient nosological safety net for the not inconsiderable number of diagnostic failures of the century. In the words of one physician: “It is yet a sufficient Answer to Patients when they enquire into their ailments to give this Return to a troublesome Enquirer, that their Disease is the Scurvy, they rest satisfied that they are devoured with a Complication of bad Symptoms . . .” (Strother 1725: 150). At the end of the century the maverick Sir John Hill declared on the title page of his *The Power of the Water Dock against the Scurvy* (1777) that “If any one is ill, and knows not his Disease, Let him suspect the Scurvy.”¹

With such confusion surrounding scurvy, one cannot with any certitude delineate the true importance

of the ailment among land dwellers in the seventeenth and eighteenth centuries, and in the absence of details of dietary patterns (and particularly of the consumption of vegetables by the population in general), there must remain considerable doubts about its true incidence. About all that is clear is that the use of the term “scurvy” was almost certainly more common than the disease itself.²

Nevertheless, epidemics did occur; Hirsch lists some 30 such outbreaks during the eighteenth century in such diverse areas as Canada, Denmark, and Russia (1885: 521–3). Among the seafaring population, however, the picture was much more clearly delineated, and right through the seventeenth and eighteenth centuries scurvy was the most feared of all the hazards associated with long sea voyages, and with good cause. As we have already noted, it has been estimated that between 1600 and 1800 scurvy accounted for some million deaths at sea and that in almost every naval campaign of any length during this period scurvy played an important role (Roddiss 1951: 72).

James Lind

The name of James Lind is deservedly associated with a substantial advance in the understanding of scurvy. A ship’s surgeon, he soon came face-to-face with the ravages of the disease among sailors. Others before him had successfully used empirically discovered remedies in the cure of scurvy, and their use of citrus fruits as the antiscorbutic remedy par excellence was well established (see, for example, Zulueta and Higuera 1981). This was presumably known to Lind, who himself quoted J. G. H. Kramer’s observation “. . . if you have oranges, lemons or citrons . . . you will, without other assistance cure this dreadful disease” (Stewart and Guthrie 1953: 154).

John Woodall in his *The Surgeons Mate* (1617) quite clearly drew attention to the prophylactic value of lemon juice – which had, in any case, already been successfully used by James Lancaster in his East Indian voyage at the beginning of the century. In 1696 William Cockburn underlined the importance of fresh fruit and vegetables to sailors in his *Sea Disease, or Treatise of their Nature, Cause and Cure* and in his *Essay on Sickness and Health* (1725). Edward Strother also had pointed out that “eating Lemons and Oranges” would cure scurvy in sailors.

But Lind’s achievement was that he subjected these empirically derived claims, and others, to the test of scientific experimentation in what has been claimed to be the first controlled clinical study in the history of medicine. Here is a description of it in Lind’s own words:

On the 20th of May, 1747, I took twelve patients in the scurvy, on board the *Salisbury* at sea. Their cases were as similar as I could have

them. They all in general had putrid gums, the spots and lassitude, with weakness of the knees. They lay together in one place, being a proper apartment for the sick in the fore-hold; and had one diet common to all, viz. water-gruel sweetened with sugar in the morning; fresh mutton-broth often times for dinner; at other times puddings, boiled biscuit with sugar etc. and for supper, barley and raisin, rice and currants, sago and wine, or the like. Two of these were ordered each a quart of cyder a day. Two others took twenty-five gutts [drops] of *elixir vitriol* three times a day, upon an empty stomach; using a garble strongly acidulated with it for their mouths. Two others took two spoonfuls of vinegar three times a day upon an empty stomach having their gruels and other food well acidulated with it, as also the gargle for the mouth. Two of the worst patients, with the tendons in the L arm rigid (a symptom none of the rest had) were put under a course of sea water. Of this they drank half a pint every day, and sometimes more or less as it operated, by way of gentle physic. Two others had each two oranges and one lemon given them each day. These they [ate] with greediness at different times, upon an empty stomach. They continued but six days under this course, having consumed the quantity that could be spared. The two remaining patients took the bigness of a nutmeg three times a day, of an electuary recommended by an hospital-surgeon, made of garlic, mustard seed, *rad. rapban.*, Balsam of Peru, and gum myrrh; using for common drink, barley water well acidulated with tamarinds; by a decoction of which with the addition of *cremor tartar*; they were gently purged three or four times during the course.

The consequence was that the most sudden and visible good effects were perceived from the use of the oranges and lemons; one of those who had taken them, being at the end of six days fit for duty. The spots were not indeed at that time quite off his body, nor his gums sound; but without any other medicine, than a gargarism of *elixir vitriol* he became quite healthy before we came into Plymouth which was on the 16th of June. The other was the best recovered of any in his condition; and being now deemed pretty well, was appointed nurse to the rest of the sick. Next to the oranges, I thought the cyder had the best effects. . . . (Stewart and Guthrie 1953: 145–7)

Lind’s note on the partial efficacy of cider has been quoted in support of the thesis that cider, as prepared traditionally in the eighteenth century, would have contained significant amounts of vitamin C and that it could have played a significant role in the pre-

vention of scurvy (French 1982: 59–66). Although this may have been the case, such evidence as is available would suggest that in general, very little of the original vitamin C would survive a fermentation procedure (Hughes 1975, 1990).

Lind's book *A Treatise of the Scurvy* appeared in 1753; a second edition was published in 1757 and a third and "updated" version in 1772 (a bicentenary reprint of the first edition was published in 1953 [Stewart and Guthrie 1953]). A French translation appeared in 1756 with subsequent reprints, and an Italian translation was published in 1766.

Not all of Lind's conclusions, however, were characterized by the same degree of scientific acumen that he displayed in his famous "scurvy trial," and it is clear that he claimed "antiscorbutic" properties for preparations that modern analysis would suggest are quite devoid of vitamin C (Hughes 1975). Nevertheless, his *experimental* demonstration that oranges and lemons could cure scurvy, his critical assessment of the literature of scurvy, his demonstration that the distinction between "land scurvy" and "sea scurvy" was a spurious one, and his advocacy of the use of various prophylactic foods by ships' crews must place him among the most significant figures in the history of nutrition.

Unfortunately, although subsequent commentators advocated the official adoption of lemon juice by the naval authorities as a prophylactic measure, they met with little success for many decades (Carpenter 1986: 73, 87). The purchase of vegetables by ship captains was sanctioned but the practice soon lapsed, and scurvy among sailors continued unabated. Indeed, when the Channel Fleet returned to port in 1780 after a cruise of only six weeks, there were 2,400 cases of scurvy present (Lloyd 1981: 12).

But in 1795, largely as a result of the advocacy of Gilbert Blane, himself a prominent and influential naval surgeon (actually "physician to the fleet"), the official issue of lemon juice to naval personnel in Britain was sanctioned (the practice was not, however, officially extended to the merchant service until 1844). Within two years, cases of scurvy in naval hospitals were rare occurrences, and there were clinicians who complained that they had never seen a true case of scurvy – a situation that would have been unthinkable a few years previously. J. Turnbull's almost monotonously successful treatment of the few scorbutic sailors arriving at Liverpool in 1848, merely providing them each with two lemons daily, was, perhaps, the final elegant vindication of Lind's pioneer study (Turnbull 1848).

Limes and Scurvy

A minor setback in keeping scurvy at a distance occurred in the 1850s when, for economic reasons, the British Admiralty contracts were changed so that the West Indian lime (*Citrus aurantiifolia*) was sub-

stituted for the lemon (*Citrus limon*). The superficial similarity between the two fruits had already led to the use of the terms "lime juicer" and "limey" by Americans in referring to British ships and sailors. Furthermore, the terms "lemon juice" and "lime juice" were used carelessly and often without distinction in referring to the two species. In fact, at the beginning of the century the lime was sometimes referred to as the "sour lemon" (Green 1824: 316).

An unfortunate consequence of this change was the reported failure of stored lime juice to offer protection against scurvy – as in the ill-fated "Nares" expedition to the Arctic in 1875. The apparent failure of lime to prevent scurvy led to the rejection by some observers of the thesis that lack of fresh fruit and vegetables was the only cause of scurvy. A. H. Smith, in a comprehensive article stimulated by outbreaks of scurvy among troops issued lime juice during the First World War, concluded that lime juice (as contrasted with lemon juice) as an antiscorbutic agent was "worthless" (Smith 1919).

The reason for this apparent lack of antiscorbutic potency on the part of lime juice preparations is still incompletely understood. It is true that lime juice has a lower vitamin C content than lemon juice, but it is still high enough to be regarded as a moderately good source of the vitamin. It is possible that the vitamin C molecule is much less stable in lime juice than in lemon juice – a belief that was apparently confirmed by Harriette Chick in 1918 using the guinea pig as an assay system. This observation would suggest that the protective and stabilizing factors present in lemon juice (high acidity, tannins, anthocyanins, etc.) are less potent in lime juice preparations. However, in a recent study, Carpenter found no essential difference between the rate of loss of ascorbic acid in lime and lemon juices (Carpenter 1991, personal communication). Whatever the true explanation, the apparent discrepant protective capacities of the two forms was an important feature in the history of scurvy, introducing a puzzling diversionary element into an otherwise fairly smooth development.

Scurvy in the Nineteenth Century

In general, the importance of scurvy diminished during the nineteenth century. It no longer posed a problem to mariners (the greater rapidity of sea vessels was a contributory factor in this respect), and it no longer occupied a position of importance in the medical texts. The eighteenth-century tendency to use the term "scorbutic" indiscriminately with reference to almost any unidentifiable or challenging condition seemed also to be dying. And an increase in the intake of vegetables (and to a lesser extent, fruit) reduced the possibility of scurvy occurring among the general population.

Nevertheless, sporadic outbreaks of scurvy occurred from time to time, and a number of these

led to detailed reports in the medical press – some of them substantial ones, such as that by J. O. Curran in 1847 of cases of scurvy following the potato famine in Ireland (Curran 1847) and that of R. Christison describing a similar outbreak in Scotland in 1846 (Christison 1847).

Scurvy outbreaks occurred most frequently in “closed” communities subjected to the same (inadequate) dietary pattern – hospitals, prisons, and workhouses. The numerous reports presented to Parliament in the nineteenth century on conditions in the prisons and workhouses provide useful information on the adequacy of the diets in these institutions (Johnston 1985). The availability of potatoes – and to a lesser extent, vegetables – was of paramount importance in this respect; fruit was almost totally absent from the normal prison and workhouse diet. Thus, the outbreaks of scurvy among convicts at Millbank in 1822 and in Pentonville in 1847 were a direct result of a lack of vegetables.

W. Baly, in 1843, in a simple type of controlled experiment, was able to show that outbreaks of scurvy at Pentonville Prison could be eliminated by including adequate amounts of potatoes in the diet (Baly 1843). In 1851, R. Boyd described two cases of scurvy (a female aged 38 and a male aged 59) in the Somerset County Pauper Lunatic Asylum resulting from “a continued diet of one meal of bread and cheese daily for three months.” Of particular significance to Boyd was the observation that this occurred “despite a plentiful allowance of cider (nearly 3 pints daily), which is supposed to be antiscorbutic” (Boyd 1851: 520; see also French 1982).

In the second half of the century potatoes were a regular feature of such institutional diets and outbreaks of scurvy diminished, although some cases still occurred among sailors despite the official adoption of lemon juice as a prophylactic. “It is very rare for London physicians to see cases of scurvy such as they are presented at our seaports,” wrote P. Black in 1876 (12), confirming Boyd’s earlier statement that “Scurvy is now seldom seen in ordinary hospital practice in this country except in the [naval] hospital ship *Dreadnought*” (Boyd 1851: 520). Nevertheless, there were still isolated epidemics in communities deprived of fresh food supplies for long periods – as among British soldiers in the Crimean War and among some of the persons on Arctic expeditions.

“Explaining” Scurvy in the Nineteenth Century

Attempts to “explain” the nature of scurvy continued. In particular, efforts were made to relate it to the absence of fresh fruit and vegetables in the diet, and by the middle of the nineteenth century there was a fair consensus that this was the main cause of the disease. Just as in the eighteenth century, when there had been attempts to accommodate scurvy within the then popular humoral theory of disease, so in the

nineteenth century there were efforts to explain it by drawing on concepts borrowed from the embryonic nutritional science of the period. This explanation involved an interesting conceptual shift from the conviction that there were positive causes of scurvy (such as coldness or too much dietary salt) to the belief that it resulted from a dietary deficiency – albeit an undefined one.

Two writers were particularly vocal in developing this theme. One was George Budd, a pioneer thinker in the development of the concept that there were clearly definable “Disorders Resulting from Defective Nutriment” (to use the title of his series of articles in the *London Medical Gazette* in 1842 [Budd 1840, 1842]). The second was Robert Barnes, who in 1864 produced his less widely available *On the Occurrence of Sea Scurvy in the Mercantile Navy*. . . . Budd, in a frequently referenced article in Tweedie’s *Library of Medicine*, stated quite unambiguously that scurvy was the result of a lack of a single essential dietary principle and of that alone; he furthermore described this antiscorbutic principle as a thermolabile one, present in fresh fruit and vegetables and with uncanny foresight prophesied that it would “be discovered by organic chemistry or the experiments of physiologists in the not too distant future” (Budd 1840: 77; Hughes 1973).

Twenty years later Barnes reaffirmed the essential dietary nature of scurvy in a 10,000-word official report. Basing his arguments on “experiments and facts” he stated, “There is no fact so well attested by the history of mankind as the dependence of scurvy upon the negation of fresh vegetable food . . . the abolition of scurvy is entirely a question of diet, and a very simple question of diet.” Barnes, basing his calculations on returns obtained from the *Dreadnought* (Naval) Hospital, attempted to arrive at a figure for the depletion period necessary for the emergence of scurvy and concluded that “symptoms such as . . . blotches like bruises, swelling of the gums, lassitude and emaciation . . . have generally been manifest in from 60 to 80 days.” He acknowledged, however, that there was a strong factor of individual variation (Barnes 1864: 330).

But what was the nature of this deficient element that was absent from scorbutogenic diets yet present in vegetables and fruits? E. A. Parkes, an eminent clinician, presented the problem in four simple sentences, logically and within the currently accepted framework of dietary knowledge in 1869:

The peculiar state of malnutrition we call scurvy is now known not to be the consequence of general starvation, though it is doubtless aided by this. Men have been fed with an amount of nitrogenous and fatty food sufficient not only to keep them in condition, but to cause them to gain weight, and yet have got scurvy. The starches also have been given in

quite sufficient amount without preventing it. It seems, indeed clear that it is to the absence of some of the constituents of the fourth dietetic group, the salts, that we must look for the cause. (1869: 492)

And the missing components in Parkes's opinion were the salts of organic acids, such as citric, malic, and tartaric - compounds known to be of vegetable origin. Parkes's advice, included in his army manual *Practical Hygiene*, was that, "In time of war every vegetable should be used which it is safe to use, and, when made into soups, almost all are tolerably pleasant to eat. . . . Good lemon juice should be issued daily (1 oz.) and it should be seen that the men take it" (1869: 494).

Parkes's theory of the nature of scurvy was by no means the only one in the field during the nineteenth century. There were many others, and as late as 1908, J. M. Cowan, writing in the standard Oxford manual *A System of Diet and Dietetics*, while admitting "the undisputed fact that a plentiful diet of fresh vegetables cures the disease" went on to state, "The exact nature of the fault is, however, undetermined" (1908: 645). Cowan referred to a number of theories of scurvy then in vogue. These were a deficiency of potash salts; a deficiency of dietary bases (alkali); ptomaine poisoning; and a specific infection (Cowan 1908: 645). Cowan's uncertainty about the cause of scurvy was echoed in 1911 by the *Encyclopedia Britannica*, which stated, ". . . the modern tendency is to suspect an unknown micro-organism . . . even among the more chemical school of pathologists it is disputed whether the cause is the absence of certain constituents in the food or the presence of some actual poison . . ." (517).

Had Cowan been aware of it, these alternative theories had already been rendered redundant by an experiment reported the previous year from Oslo - in Carpenter's opinion "the most important single paper in the whole history of this subject" (Carpenter 1986: 173). This was the classic study by A. Holst, in which he produced scurvy in guinea pigs by feeding them a grain diet - in an experiment originally designed to study the nature of beriberi.

Holst had fortuitously chosen one of the few species unable to produce vitamin C endogenously and had fed them a diet completely lacking in vitamin C. The guinea pigs died after a month on the diet, and it was noted that the tissues showed degenerative changes similar to those known to occur in human scurvy (Holst and Frohlich 1907). Until the characterization of the "antiscorbutic factor" and the development of chemical methods of assessing it in the 1930s, the prevention of scurvy in guinea pigs was the standard technique for measuring the antiscorbutic potency of a preparation.

The last quarter of the nineteenth century also witnessed a proliferation of interest, both in America and

in Europe, in "infantile scurvy" - a form of avitaminosis C occurring in very young children and characterized primarily by defective bone development. It was later referred to as "Barlow's disease," after Thomas Barlow, a London physician, who described its true nature and distinguished it from rickets, with which it was frequently associated and often confused (Evans 1983).

It is interesting to note that the emergence of infantile scurvy, like that of "sailors' scurvy" three centuries previously, was probably a consequence of developments in technology, namely, too great a dependence on processed foodstuffs from which the vitamin C had been destroyed and, in particular, evaporated and condensed milks, the use of which increased significantly during the second half of the nineteenth century (Carpenter 1986: 158-72). The subsequent introduction of orange juice supplements for babies ensured the sudden eradication of "Barlow's disease," although isolated reports from Germany, as late as the 1980s, described its occurrence in babies fed only on oat gruel.

Experimental Scurvy in Humans

Following the identification of ascorbic acid as the "antiscorbutic factor," human volunteers attempted to induce scurvy in themselves by subsisting on a vitamin C-free diet. J. Menshing, in Germany, ate a virtually vitamin C-free diet for 100 days; his blood vitamin C fell to zero but there were no discernible signs of scurvy. H. Rietschel and H. Schick repeated the experiment, extending the period of deprivation to 160 days but again without the appearance of any overt signs of scurvy (Rietschel and Mensching 1939). J. H. Crandon, in the United States, deprived himself of vitamin C and noted the appearance of the first recognizable signs of scurvy after some 5 months (Crandon, Lund, and Dill 1940).

However, the first controlled study designed to produce scurvy in humans was undertaken at Sheffield in England during the latter stages of the Second World War and published in 1953 (eight years after the conclusion of the work) as Medical Research Council Special Report (Medical Research Council 1953). Twenty young volunteers (aged 21 to 34) took part in the experiment; anecdotal reports indicate that they were conscientious objectors to military service, although, understandably perhaps, this information was not included in the official report of the study.

All participants were placed on a vitamin C-free diet. Ten received no supplements; seven were given 10 mg ascorbic acid daily; and three received 70 mg ascorbic acid daily. The general picture to emerge was that the classical signs of scurvy (hyperkeratotic follicles, gum lesions, impaired wound healing - in that order) were experienced by those in the deficient group but not by either of the supplemented groups.

The hair follicle changes occurred after 17 weeks of deprivation and the gum changes after 26 weeks – somewhat later than the corresponding period calculated by Barnes in 1864. The difference presumably reflected differences in the subjects' vitamin C status at the beginning of the deprivation period. Furthermore, a supplement of 10 mg ascorbic acid was sufficient to restore the scorbutic subjects to normal health.

The report recommended that “in order to arrive at a figure for a daily allowance which covers individual variations and includes a safety margin, it is suggested that the minimum protective dose of 10 mg be trebled” (Medical Research Council 1953: 145). For 40 years the United Kingdom Recommended Daily Allowance for vitamin C remained at 30 mg, and the experimental evidence for this amount was the single “Sheffield Study” of 1944 – a study that, in terms of sex and age alone, it may be argued, was based on a completely unrepresentative population group.

A similar investigation, but with fewer participants, was reported from the United States in 1969. The subjects in this instance were prisoners who “bought” remission periods by subjecting themselves to scurvy. The signs of deficiency appeared somewhat earlier than in the “Sheffield Study” – skin lesions in 8 to 13 weeks and gum changes in 11 to 19 weeks (Hodges et al. 1971). Both the Sheffield and the U.S. projects included wide-ranging physiological and pathological examinations, and further projects of this nature are unlikely to add very much to our clinical knowledge of the condition.

There remain, of course, peripheral problems that are of some academic interest but of little practical significance, such as the identification of further species unable to synthesize ascorbic acid and, therefore, susceptible to scurvy – and the mirror image of this, the search for a mutant form of guinea pigs resistant to scurvy. One of the more interesting current problems centers on the apparent refractiveness of the elderly to scurvy; it has been shown that persons over 80 years of age may have extremely low blood levels of ascorbic acid without the emergence of any of the expected clinical features (Hughes 1981: 60).

Scurvy Today

Our current knowledge of the vitamin C requirement of the body allows us, in a perverse way, to assess the validity of some historical claims of dietary interest. L. Cornaro, who reputedly attained a century of life by prudent living, recommended for the poorer and older members of the community a diet of bread, panada, and eggs. A diet so structured, however, would be almost completely lacking in vitamin C and almost certain to produce scurvy in a matter of months (Cornaro 1768: 98). Similarly, the numerous accounts of survival for considerable periods on diets, which simple inspection indicates to be completely

vitamin C-free (such as that of John Ferguson, who in the eighteenth century lived, allegedly for 18 years, only on water, whey, and barley water [Umfreville 1743]), must now be regarded as largely apocryphal.

The life, and untimely death, of William Stark was more in keeping with what we now know of scurvy and vitamin C. Stark, perhaps the best known of all dietary masochists, achieved historical immortality by experimenting on his own body. He may be regarded as the founder of the now firmly rooted “Death by Diet” brigade. A 29-year-old unmarried physician living in London, he decided to test whether very simple diets would support health. From July 1769 to February 1770 he subsisted on diets such as meat, bread, and water; or bread and cheese; or honey and bread; or sugar and water. This highly scorbutigenic regime was relieved only by a small amount of fruit on one occasion in December 1769. Stark died on February 23, 1770, almost certainly from scurvy, judging by the entries in his detailed diary of the study. It is interesting to note, despite his general lassitude and fatigue, entries such as “had strong desires,” “Venus semel,” and “Venus bias,” which appeared at least weekly in his diary up to within a fortnight of his death, suggesting little impairment of the procreative capacity even in severe scurvy (Smyth 1788; Drummond and Wilbraham 1935).

Scurvy is today a rare condition – so rare that individual cases usually merit a short note in the medical press. Persons on a mixed, balanced diet would normally be expected to daily ingest more than the amount of ascorbic acid required to protect them against the disease. Thus, in the United Kingdom the mean ascorbic acid intake is some 60 mg daily – 50 percent above the current U.K. Reference Daily Intake of 40 mg. The only persons likely to fall into a scurvy-risk category are those who, for economic or dietary reasons, subsist on a diet deficient in fruit, vegetables, and fresh meat – such as a diet based primarily on nuts and grain and dairy products or the proverbial American scorbutigenic diet of doughnuts and black coffee.

Indeed, even the ingestion of large amounts of fresh meat (particularly liver) will provide sufficient amounts of vitamin C to prevent scurvy in the comparative absence of fruit and vegetables. This explains why Eskimos (the name of which, apparently, means “raw flesh eaters”) remain scurvy-free even though their intake of plant sources is minimal. A turn-of-the-century account underlines this fact:

In 1893, at Kharborova, a Samoyad settlement on the Yugor Straits . . . six Russian priests, whose religion forbade them to eat reindeer or other such meats, but allowed salted fish, were left in a hut by a wealthy mine-owner to pass the winter. . . . A small Russian peasant boy was left to wait on them. The priests lived almost exclusively on tea, bread and salted fish; the boy

lived upon similar food, except that instead of the salted fish he ate fresh reindeer meat. None of them had any vegetables. In the following May, when the Samoyads and peasant traders returned, they found that all the six priests had died of scurvy, whereas the little boy, who had lived upon fresh meat and had not eaten salted fish was alive and well, and had buried all his late masters in the snow. . . . (Jackson and Harley 1900: 252-3)

Recently reported cases of scurvy include that of a 9-year-old American girl who consumed nothing but tuna sandwiches and iced tea (without lemon) (Ellis, Vanderveen, and Rasmussen 1984), and that of a 24-year-old British male whose main form of sustenance was peanut butter sandwiches (Pozniak and Tomson 1983).

Epidemics of scurvy have been equally rare in the twentieth century. They occurred on a limited scale among soldiers in World War I (Smith 1919) and, particularly, among Indian personnel fighting in Iraq, where some 11,000 cases were reported (Willcox 1920). J. D. Comrie stated that he had personally examined 600 cases of scurvy in 1919 in North Russia (Comrie 1920).

Even today, however, we have the occasional reminder that the history of scurvy is not as completely closed a book as modern science would perhaps lead us to believe. Recently it has been reported that scurvy remains a serious public-health problem for Ethiopian refugees in Sudan and Somalia, where the incidence of scurvy in six camps ranged from 14 percent to 44 percent. It was found that the standard relief food (mainly cereals, legumes, and oil) distributed to the refugees was almost completely deficient in vitamin C (Desenclos et al. 1989).

The condition sometimes described as "chronic scurvy" or hypovitaminosis C (in which a person subsists on a suboptimal intake of vitamin C but without the emergence of the clinical features of overt scurvy) may affect substantial sectors of a population and, in particular, the institutionalized elderly; but its clinical significance, if any, is a matter of continuing dispute that need not concern us here.

R. E. Hughes

Notes

1. The water dock was not the only plant to have been the subject of a monograph devoted to its supposed antiscorbutic properties; Andreas Moellenbrok's *Cochlearia Curiosa* was translated into English in 1676, and a century later John Edwards published his . . . *Goose Grass or Clivers and its Efficacy in the Cure of the most inveterate Scurvy*.
2. Ironically, the situation is today completely reversed. In a recent paper it was pointed out that "Scurvy is a disease which can mimic other more serious disorders such as deep

vein thrombosis and systemic bleeding disorders . . . [and] because the clinical features of scurvy are no longer well appreciated scorbutic patients are often extensively evaluated for other disorders" (Reuler, Brondy, and Cooney 1985).

Bibliography

- Baly, W. 1843. On the prevention of scurvy in prisoners, pauper [and] lunatic asylums etc. *London Medical Gazette* 1: 699-703.
- Barnes, R. 1864. Report on the occurrence of sea scurvy . . . *Public Records (Commissioners Reports)* 28.
- Black, P. 1876. *Scurvy in high latitudes*. London.
- Boyd, R. 1851. Two cases of scurvy. *Lancet* 1: 519-20.
- Budd, George. 1840. Scurvy. In *The library of medicine*, Vol. 5, ed. A. Tweedie, 58-95. London.
1842. Disorders resulting from defective nutriment. *London Medical Gazette* 2: 632-6, 712-16, 743-9, 906-15.
- Carpenter, K. J. 1986. *The history of scurvy and vitamin C*. Cambridge.
- Christison, R. 1847. Account of scurvy as it has lately appeared in Edinburgh . . . *Monthly Journal of Medical Science* 8: 1-22.
- Clay, R. M. 1909. *The medieval hospitals of England*. London.
- Comrie, J. D. 1920. Scurvy in North Russia. *Edinburgh Medical Journal* 24: 207-15.
- Cornaro, L. 1768. *Discourses on a sober and temperate life*. London.
- Counsell, J. N., and D. H. Hornig, eds. 1981. *Vitamin C (ascorbic acid)*. London and Englewood Cliffs, N.J.
- Cowan, J. M. 1908. Diet in diseases of the blood. In *A system of diet and dietetics*, ed. G. A. Sutherland, 625-52. London.
- Crandon, J. H., C. C. Lund, and D. B. Dill. 1940. Human experimental scurvy. *New England Journal of Medicine* 223: 353-69.
- Curran, J. O. 1847. Observations on scurvy as it has lately appeared throughout Ireland and in several parts of Great Britain. *Dublin Quarterly Journal of Medical Science* 4: 83-134.
- Desenclos, J. C., A. M. Berry, R. Padt, et al. 1989. Epidemiological patterns of scurvy among Ethiopian refugees. *Bulletin of the World Health Organization* 66: 309-16.
- Dodoens, R. 1578. *A nieuwe herball or historie of plantes* . . . translated by Henry Lyte. London.
- Drummond, J. C., and A. Wilbraham. 1935. William Stark, M.D. 1964. *The Englishman's food*. Revised edition. London.
- Ellis, C. N., E. E. Vanderveen, and J. E. Rasmussen. 1984. Scurvy. A case caused by peculiar dietary habits. *Archives of Dermatology* 120: 1212-14.
- The Encyclopaedia Britannica*. 1911. Vol. 24, 517. Cambridge.
- Englard, S., and S. Seifter. 1986. The biochemical functions of ascorbic acid. *Annual Review of Nutrition* 6: 365-406.
- Evans, P. D. 1983. Infantile scurvy: The centenary of Barlow's disease. *British Medical Journal* 287: 1862-3.
- Floyer, John. 1722. *History of cold-bathing*. Fifth edition. London.
- French, R. K. 1982. *The history and virtues of cyder*. New York and London.
- Gerarde, J. 1633. *The herball or generall historie of plantes* . . . amended by Thomas Johnson. London.
- Green, T. [1824]. *The universal herball*. London.
- Harvey, G. 1675. *The disease of London or a new discovery of the scorvey*. London.
- Henslow, G. 1899. *Medical works of the fourteenth century*. London.

- Hess, A. S. 1920. *Scurvy past and present*. London.
- Hirsch, A. 1883–6. *Handbook of geographical and historical pathology*; Vol. 2, trans. C. Creighton. London.
- Hodges, R. E., J. Hood, J. E. Canham, et al. 1971. Clinical manifestations of ascorbic acid deficiency in man. *American Journal of Clinical Nutrition* 24: 432–43.
- Holst, A., and T. Frohlich. 1907. Experimental studies relating to ship beri-beri and scurvy. II. On the etiology of scurvy. *Journal of Hygiene* 7: 634–71.
- Hughes, R. E. 1973. George Budd (1808–1882) and nutritional deficiency diseases. *Medical History* 17: 127–35.
1975. James Lind and the cure of scurvy. *Medical History* 19: 342–51.
1981. *Vitamin C*. London.
1990. The rise and fall of the “antiscorbutics”; some notes on the traditional cures for “land scurvy.” *Medical History* 34: 52–64.
- In press. De Scelytyrbe . . . or a Traetise of the Scorbut. *The Journal of Welsb Ecclesiastical History*.
- Jackson, F. G., and V. Harley. 1900. An experimental inquiry into scurvy. *Proceedings of the Royal Society* 66: 250–64.
- Johnston, V. J. 1985. *Diet in workhouses and prisons 1835–1895*. New York and London.
- King, L. S. 1970. *The road to medical enlightenment*. London and New York.
- Letheby, H. 1870. *Lectures on food*. London.
- Lind, J. 1753. (See Stewart and Guthrie 1953.)
- Lloyd, C. C. 1981. In *Starving sailors: The influence of nutrition upon naval and maritime history*; ed. J. Watt, E. J. Freeman, and W. F. Bynum, 9–15. Bristol, England.
- Major, R. H. 1978. *Classic descriptions of disease*. Third edition. Springfield, Ill.
- Maynwaringe, Everard. 1672. *Marbus Polyrbizos and Polymorphaeus: A treatise of the scurvy*. . . . London.
- Medical Research Council. 1953. *Vitamin C requirements of human adults*. Special Report Series, No. 280. London.
- Moffet, T. 1746. *Health's improvement*. London.
- Parkes, E. A. 1869. *A manual of practical hygiene*. London.
- Powicke, F., ed. 1926. Richard Baxter, The poor husbandman's advocate. *Bulletin of the John Rylands Library* 10: 464.
- Pozniak, A. L., and C. R. Tomson. 1983. Clinical Curio: Scurvy in a 24-year-old. *British Medical Journal* 286: 881.
- Prentice, E. P. 1939. *Hunger and history*. New York and London.
- Reuler, J. B., V. C. Brondy, and T. G. Cooney. 1985. Adult scurvy. *Journal of the American Medical Association* 253: 805–7.
- Rietschel, H., and J. Mensching. 1939. Experimenteller C-Vitamin-hunger am Menschen, ein Betrag zur Frage des C-Vitaminbedarfs. *Klinische Wochenschrift* 18: 273.
- Riverius, Lazarus. (Rivièr, Lazare) 1655. *The practice of physick in several books . . . by Nicholas Culpeper . . . Abdiab Cole . . . William Rowland being chief of translation of the works of . . . Lazarus Riverius*. London.
- Roddis, L. S. 1951. *James Lind*. London.
- Salaman, R. 1970. *The history and social influence of the potato*. Cambridge.
- Salmon, William. 1696. *The family-dictionary*. Second edition. London.
- Shapter, Thomas. 1847. “On the recent occurrence of Scurvy in zar and the neighbourhood.” *Medical Gazette* May 21: 945–8.
- Smith, A. H. 1919. A historical inquiry into the efficacy of lime-juice for the prevention and cure of scurvy. *Journal of the Royal Army Medical Corps* 32: 93–116, 188–208.
- Smyth, J. C. 1788. *The works of the late William Stark, M.D. with experiments dietetical and statical, revised and published by James Carmichael Smyth, M.D., F.R.S.* London.
- Society of Gentlemen. 1777. Potatoes. *The complete farmer*. Third edition. London.
- Stephens, W. 1732. *Dolaeus upon the cure of the gout by milk-diet*. London.
- Stewart, C. P., and D. Guthrie, eds. 1953. *Lind's treatise on scurvy*. Edinburgh.
- Strother, Edward. 1725. *An essay on sickness and health*. London.
- Sutherland, G. A., ed. 1908. *A system of diet and dietetics*. London.
- Turnbull, J. 1848. Observations on the scurvy. *Lancet* 1: 601–4.
- Turner, W. 1568. *The first and second parts of the herbal . . .* London.
- Umfreville, T. 1743. *The case of Mr. John Ferguson . . .* London.
- Vaughan, William. 1630. *The newlanders cure . . .* London.
- [Wier, J.] 1622. *Discourse of the scorby, translated out of Wyers' observations*. London.
- Willcox, W. H. 1920. The treatment and management of diseases due to deficiency of diet: Scurvy and beri-beri. *British Medical Journal* 1: 73–7.
- Woodall, J. 1617. *The surgions mate or military and domestic surgery*. London.
- Young, Arthur. 1780. *A tour in Ireland*, Vol. 2, Second edition. London.
- Zulueta, J. de. and L. Higuera. 1981. Health and navigation in the South Seas: The Spanish experience. In *Starving sailors: The influence of nutrition upon naval and maritime history*; ed. J. Watt, E. J. Freeman, and W. F. Bynum, 85–99. Bristol, England.



IV.E

Food-Related Disorders

IV.E.1 Anorexia Nervosa

Definition

Anorexia nervosa is a psychophysiological disorder – usually of young females – characterized by a prolonged refusal to eat or maintain normal body weight, an intense fear of becoming obese, a disturbed body image in which the emaciated patient feels overweight, and the absence of any physical illness that would account for extreme weight loss. The term “anorexia” is actually a misnomer, because genuine loss of appetite is rare and usually does not occur until late in the illness. In reality, most anorectics are obsessed with food and constantly struggle to deny natural hunger.

In anorexia nervosa, normal dieting escalates into a preoccupation with being thin, profound changes in eating patterns, and a loss of at least 25 percent of the original body weight. Weight loss is usually accomplished by a severe restriction of caloric intake, with patients subsisting on fewer than 600 calories per day. Contemporary anorectics may couple fasting with self-induced vomiting, use of laxatives and diuretics, and strenuous exercise.

The most consistent medical consequences of anorexia nervosa are amenorrhea (ceasing or irregularity of menstruation) and estrogen deficiency. In most cases amenorrhea follows weight loss, but it is not unusual for amenorrhea to appear before noticeable weight loss has occurred. The decrease in estrogen causes many anorectics to develop osteoporosis, a loss of bone density that is usually seen only in postmenopausal women (Garfinkel and Garner 1982).

By the time the anorectic is profoundly underweight, other physical complications resulting from severe malnutrition begin to appear. These include bradycardia (slowing of the heartbeat), hypotension (loss of normal blood pressure), lethargy, hypothermia, constipation, the appearance of “lanugo” or fine

silky hair covering the body, and a variety of other metabolic and systemic changes.

In addition to the physical symptoms associated with chronic starvation, anorectics also display a relatively consistent cluster of emotional and behavioral characteristics, the most prominent of which grows out of the anorectic's deviation from normal eating habits. Severe restriction of food intake is sometimes alternated with bulimic phases, in which the anorectic engages in uncontrolled or excessive eating followed by self-induced vomiting and laxative abuse. Other unusual eating habits may include monotonous or eccentric diets, hoarding or hiding of food, and obsessive preoccupation with food and cooking for others.

Emotionally, anorexic patients are often described as being perfectionist, dependent, introverted, and overly compliant. Although studies have failed to find a consistent psychiatric symptom pattern for the disorder, frequently reported neurotic traits include obsessive-compulsive, hysterical, hypochondriacal, and depressive symptoms. A decrease in or disappearance of sexual interest is also a frequent concomitant of anorexia nervosa.

A distorted body image is an almost universal characteristic of anorectics, with many patients insisting they are overweight even when their bodies are extremely emaciated. As a result, most individuals with anorexia nervosa deny or minimize the severity of their illness and are usually highly resistant to therapy. The anorectic's refusal to acknowledge her nutritional needs, and her steadfast insistence that nothing is wrong, make anorexia nervosa one of the most recalcitrant disorders in contemporary medicine.

Distribution and Incidence

Once considered to be extremely rare, the reported incidence of anorexia nervosa has more than doubled during the past 20 years (Herzog and Copeland 1985). The disorder is especially prevalent among adolescent and young adult women. Ninety to 95 percent

of anorectics are young and female, and as many as 1 in 250 females between 12 and 18 years of age may develop the disorder. The exact incidence of anorexia and other eating disorders is difficult to determine, however, because of problems in conducting reliable epidemiological studies and the small samples on which many such studies are based (Crisp, Palmer, and Kalucy 1976).

The onset of anorexia nervosa occurs most often during adolescence, although some patients have become anorexic as early as age 11 and as late as the sixth decade of life. Patients are typically high achievers, with normal or above average intelligence. They also tend to come from middle- or upper-class families, although evidence of anorexia nervosa among working-class and poverty-class women is growing (Bulik 1987; Gowers and McMahon 1989; Dolan 1991).

Anorexia nervosa is comparatively rare in men: Approximately 5 to 10 percent of anorectics are male. The clinical picture for male anorexic patients is also much different from that for women. In general, male anorectics tend to display a greater degree of psychopathology, are often massively obese before acquiring the disorder, are less likely to be affluent, and are even more resistant to therapy than their female counterparts (Garfinkel and Garner 1982). There is growing evidence, however, that anorexia nervosa and bulimia are more common among men than previously believed. This is particularly true among homosexual men, who tend to experience more dissatisfaction with body image than do heterosexual men (Yager et al. 1988; Silberstein et al. 1989; Striegel-Moore, Tucker, and Hsu 1990).

Anorexia nervosa was also once thought to be comparatively rare among American blacks, Hispanics, Native Americans, lesbians, first- and second-generation ethnic immigrants, and individuals from disadvantaged socioeconomic backgrounds (Herzog and Copeland 1985). Recent research in this area, however, has indicated that the incidence of anorexia and other eating disorders among these groups is much higher than previously thought (see the section "Geographic and Demographic Features").

Etiology and Epidemiology

Although the etiology of anorexia nervosa is an area of intense investigation, researchers have yet to reach a consensus about the origin of the disorder. The most sophisticated thinking on the subject regards anorexia nervosa as a multidetermined disorder that involves an interplay of biological, psychological, and cultural factors. Advocates of this model view these three etiological factors as reciprocal and interactive and believe it is simplistic to isolate one component as the single underlying cause of the disorder (Garfinkel and Garner 1982; Brumberg 1988).

Joan Brumberg (1988) has developed a multideter-

mined etiological model based on a two-staged conceptualization of anorexia nervosa that delineates the relative impact of sociocultural influences and individual biological and psychological variables in precipitating the disorder. In the first stage – the "recruitment" phase of the illness – sociocultural factors play the dominant role. During this period, cultural assumptions that associate thinness with female beauty lead certain women into a pattern of chronic dieting. Indeed, research on the sociocultural causes of anorexia nervosa has linked the increased incidence of anorexia nervosa and other eating disorders with the tremendous cultural attention given to dieting and food, increasingly thinner standards of beauty, and the fitness movement (Schwartz, Thompson, and Johnson 1982; Chernin 1985; Orbach 1986; Bordo 1993). Yet sociocultural variables alone cannot explain why some women but not others move from chronic dieting to anorexia nervosa. Therefore, other individual factors must be implicated in the final development of the illness.

Brumberg's model of anorexia nervosa relies on a second stage – career or acclimation – to correct the shortcomings of sociocultural explanations of the disorder. During the career phase, specific biological and psychological features determine which individuals develop the full-blown psychopathology of anorexia nervosa. In order to explain the transition between the recruitment and career phases of anorexia nervosa, Brumberg relies on recent research in the biological and social sciences, which has sought to uncover the unique physiological and psychological characteristics of anorexic patients.

Since the early 1900s, a number of different endocrinological and neurological abnormalities have been postulated as underlying biological causes of anorexia nervosa: hormonal imbalance, dysfunction in the satiety center of the hypothalamus, lesions in the limbic system of the brain, and irregular output of vasopressin and gonadotropin (Herzog and Copeland 1985). The search for a biomedical cause of anorexia nervosa is made difficult, however, by the fact that chronic starvation itself produces extensive changes in hypothalamic and metabolic function. Researchers in this area have yet to find a common biological characteristic of the anorexic population that is unmistakably a cause rather than a consequence of extreme weight loss and malnutrition (Brumberg 1988).

A more satisfactory explanation of the biological factors that contribute to the "career" phase of anorexia nervosa is the "addiction to starvation" model proffered by the British psychiatrists George I. Szmukler and Digby Tantum (1984). According to Szmukler and Tantum, patients with fully developed anorexia nervosa are physically and psychologically dependent on the state of starvation. Much like alcoholics and other substance abusers, anorectics find something gratifying or tension-relieving about the state of starvation and possess a specific physiological substrate that

makes them more susceptible to starvation dependence than individuals who merely engage in chronic dieting. Szmukler and Tantum add, however, that starvation dependence is not the total explanation of anorexia nervosa. Rather, they believe that starvation dependence acts in conjunction with a range of sociocultural, psychological, and familial factors that encourage certain individuals to use anorexic behavior as a means of expressing personal anguish.

Current psychological models of anorexia nervosa fall into three basic categories: psychoanalytic, family systems, and social psychology. In both the psychoanalytic and family systems models, anorexia nervosa is seen as a pathological response to the developmental crisis of adolescence. Orthodox psychoanalysts, drawing on the work of Sigmund Freud, view the anorectic as a girl who fears adult womanhood and who associates eating with oral impregnation (Brumberg 1988). Family systems theory, however, offers a more complex explanation of the relationship between adolescence and anorexia nervosa. On the basis of clinical work with anorectics and their families, family systems therapists have found that the majority of anorexic patients are "enmeshed," meaning that the normal process of individuation is blocked by extreme parental overprotectiveness, control, and rigidity. Anorexia nervosa is therefore seen as a form of adolescent rebellion against parental authority (Minuchin, Rosman, and Baker 1978; Bruch 1988).

Research in social psychology and the field of personality has devised several psychological tests to distinguish the psychological characteristics of anorectics from others in their age-group. One study has shown that although many of the psychological traits of anorectics and other women are indistinguishable, anorectics display a markedly higher degree of ineffectiveness and lower amount of self-esteem. Other studies have proposed that anorectics have actual cognitive problems with body imaging; still others suggest a relationship between anorexia nervosa and sex-role socialization (Garfinkel and Garner 1982).

Some researchers have attempted to fit anorexia nervosa within other established psychiatric categories such as affective disorders and obsessional neurosis. Many anorectics, in fact, display behavior patterns associated with obsessive-compulsive disorders: perfectionism, excessive orderliness and cleanliness, meticulous attention to detail, and self-righteousness. This correlation has led a number of researchers to suggest that anorexia nervosa is itself a form of obsessive-compulsive behavior (Rothenberg 1986). Depressive symptoms are also commonly seen in many patients with anorexia nervosa. Various family, genetic, and endocrine studies have found a correlation between eating disorders and depression. However, the association between anorexia nervosa and other psychiatric disorders remains controversial (Garfinkel and Garner 1982; Herzog and Copeland 1985).

History

The extent to which people suffered from anorexia nervosa in the past has been a subject of much historical debate. Some clinicians and medical historians have suggested that anorexia nervosa was first identified in 1689 by the British doctor Richard Morton, physician to James II (Bliss and Branch 1960; Silverman 1983). The medieval historian Rudolph Bell (1985) has dated the origins of anorexia nervosa even earlier, claiming that the medieval female saints, who were reputed to live without eating anything except the Eucharist, actually suffered from anorexia nervosa.

Other historians, however, have argued that attempts to label all historical instances of food refusal and appetite loss as anorexia nervosa are simplistic and maintain that the historical record is insufficient to make conclusive diagnoses of individual cases (Bynum 1987; Brumberg 1988). Although these historians agree that the final physiological stage of acute starvation may be the same in contemporary anorectics and medieval ascetics, the cultural and psychological reasons behind the refusal to eat are quite different. Thus, to reduce both to a single biomedical cause is to overlook the variety of social and cultural contexts in which certain individuals have chosen to refuse food.

The modern disease classification of anorexia nervosa emerged during the 1860s and 1870s, when the work of public asylum keepers, elite British physicians, and early French neurologists partially distinguished anorexia nervosa from other diseases involving loss of appetite (Brumberg 1988). In 1859, the American asylum physician William Stout Chipley published the first American description of *sitomania*, a type of insanity characterized by an intense dread or loathing of food (Chipley 1859). Although Chipley found *sitophobia* in patients from a broad range of social groups and age-groups, he identified a special form of the disease that afflicted adolescent girls.

Chipley's work was ignored by his contemporaries, however, and it was not until the 1870s, when two influential case studies by the British physician William Withey Gull and the French alienist Charles Lasègue (Lasègue 1873; Gull 1874) were published, that physicians began to pay significant attention to anorexia in girlhood. Gull's primary accomplishment was to name and establish anorexia nervosa as a coherent disease entity, distinct from mental illnesses in which appetite loss was a secondary feature and from physical "wasting" diseases such as tuberculosis, diabetes, or cancer. Despite widespread acclaim for Gull's work with anorexic patients, however, late-nineteenth-century clinicians generally rejected the conception of anorexia nervosa as an independent disease. Instead, they viewed it either as a variant of hysteria that affected the gastrointestinal system or as a juvenile form of neurasthenia (Brumberg 1988).

Nineteenth-century physicians also tended to

focus on the physical symptom of not eating and ignored the anorexic patient's psychological reasons for refusing food. An important exception was Lasègue, who was the first to suggest the significance of family dynamics in the genesis and perpetuation of anorexia nervosa. Because of the somatic emphasis of nineteenth-century medicine, however, most medical practitioners of that time disregarded Lasègue's therapeutic perspective. Instead, they directed medical intervention toward restoring the anorectic to a reasonable weight and pattern of eating rather than exploring the underlying emotional causes of the patient's alleged lack of appetite (Brumberg 1988).

In the twentieth century, the treatment of anorexia nervosa changed to incorporate new developments within medical and psychiatric practice. Before the Second World War, two distinct and isolated models dominated medical thinking on anorexia nervosa. The first approach was rooted in late-nineteenth-century research in organotherapy, a form of treatment based on the principle that disease resulted from the removal or dysfunction of secreting organs and glands (Brumberg 1988). Between 1900 and 1940, a variety of different endocrinologic deficiencies were proposed as the cause of anorexia nervosa. In 1914, Morris Simmonds, a pathologist at the University of Hamburg, published a clinical description of an extreme cachexia due to destruction of the anterior lobe of the pituitary. Because patients with anorexia nervosa and those with Simmonds's disease shared a set of common symptoms, many clinicians assumed that a deficiency in pituitary hormone was the cause of both conditions (Brumberg 1988).

Other researchers implicated thyroid insufficiency as the cause of anorexia nervosa. Research conducted at the Mayo Clinic in Rochester, Minnesota, during the period between the two world wars established the relationship between thyroid function and body weight and led many physicians to regard anorexia nervosa as a metabolic disorder caused by a deficiency in thyroid hormone. Throughout the 1920s and 1930s, insulin, antuitrin, estrogen, and a host of other hormones were also employed in the treatment of anorexia nervosa (Brumberg 1988).

The second major approach to anorexia nervosa in the early twentieth century grew out of the field of dynamic psychiatry, which emerged during the 1890s and early 1900s. Beginning in the last decade of the nineteenth century, practitioners in dynamic psychiatry increasingly focused on the life history of individual patients and the emotional sources of nervous disease. Two of the leading pioneers in this new field – Sigmund Freud and Pierre Janet – were the first to suggestively link the etiology of anorexia nervosa with the issue of psychosexual development. According to Freud, all appetites were expressions of libido or sexual drive. Thus, not eating represented a repression of normal sexual appetite (Freud 1959). Similarly, Janet asserted that anorexic girls refused food in

order to retard normal sexual development and forestall adult sexuality (Janet 1903).

Because of the enormous popularity of endocrinologic explanations, the idea of anorexia nervosa as a psychosexual disturbance was generally overlooked for more than 30 years. By the 1930s, however, the failure of endocrinologic models to establish either a predictable cure or a definitive cause of anorexia nervosa, the growing reputation of the Freudian psychoanalytic movement, and increased attention to the role of emotions in disease led a number of practitioners to assert the value and importance of psychotherapy in the treatment of anorexia nervosa. Although biomedical treatment of the disorder continued, most clinicians argued that successful, permanent recovery depended on uncovering the psychological basis for the anorectic's behavior. Following up on the work of Freud and Janet, orthodox psychiatrists during this time postulated that refusal to eat was related to suppression of the sexual appetite and claimed that anorexic women regarded eating as oral impregnation and obesity as pregnancy (Brumberg 1988).

After World War II, a new psychiatric view of eating disorders, shaped largely by the work of Hilde Bruch, encouraged a more complex interpretation of the psychological underpinnings of anorexia nervosa. Although Bruch agreed that the anorectic was unprepared to cope with the psychological and social consequences of adulthood and sexuality, she also stressed the importance of individual personality formation and factors within the family that contributed to the psychogenesis of anorexia nervosa. Here, Bruch revived Lasègue's work on the role of family dynamics in anorexia nervosa. According to Bruch, the families of most anorexic patients were engaged in a dysfunctional style of familial interaction known as "enmeshment": Such families are characterized by extreme parental overprotectiveness, lack of privacy of individual members, and reluctance or inability to confront intrafamilial conflicts. Although superficially these families appeared to be congenial, Bruch wrote, this harmony was achieved through excessive conformity on the part of the child, which undermined the child's development of an autonomous self. Anorexia nervosa, according to Bruch, was therefore a young woman's attempt to exert control and self-direction within a family environment in which she otherwise felt powerless (Bruch 1973, 1988).

Bruch was also primarily responsible for the tremendous growth in the popular awareness of anorexia nervosa and other eating disorders in the 1970s and 1980s. Through her book, *The Golden Cage: The Enigma of Anorexia Nervosa* (1978), which sold over 150,000 copies, and numerous articles in *Family Circle* and other popular magazines, Bruch brought anorexia nervosa into common American parlance.

At the same time that the American public was becoming increasingly aware of anorexia nervosa, the number of reported cases of the disorder grew

tremendously. This phenomenon has led some clinicians and social commentators to suggest that the popularization process itself may promote a "sympathetic host environment" for the disorder (Striegel-Moore, Silberstein, and Rodin 1986). As Bruch herself observed: "Once the discovery of isolated tormented women, it [anorexia nervosa] has now acquired a fashionable reputation, of being something to be competitive about. . . . This is a far cry from the twenty-years-ago anorexic whose goal was to be unique and suggests that social factors may impact the prevalence of the disorder" (Bruch 1988: 3-4).

Geographic and Demographic Features

Until recently, anorexia nervosa was believed to be a disorder largely confined to the United States, Canada, and Western Europe. Researchers also thought that the disease was virtually nonexistent in people of color and/or those from disadvantaged socioeconomic backgrounds. As early as 1880, S. Fenwick observed that anorexia nervosa "is much more common in the wealthier class of society than amongst those who have to procure their bread by daily labor" (Fenwick 1880: 11). This image of anorexia nervosa as a disease of abundance has persisted into the present day. Many researchers suggest that individuals from non-Western societies, minority groups, and impoverished backgrounds are "protected" from eating disorders because thinness is not highly valued in these communities, and fatness is often viewed as a sign of health and prosperity (Andersen and Hay 1985; Gray, Ford, and Kelly 1987; Gowers and McMahon 1989). The apparent absence of the disorder in developing nations and its high incidence among affluent social groups in Westernized countries led many clinicians to classify anorexia nervosa as a "culture bound" syndrome, meaning a disorder that is restricted to certain cultures primarily because of their distinctive psychosocial features (Prince 1985).

As a result of these views, none of the early literature on anorexia nervosa mentioned individuals from minority groups, non-Western countries, or lower socioeconomic classes (Bruch 1966; Kendall et al. 1973; Garfinkel and Garner 1980). The first cases of nonwhite anorectics appeared in a paper by M. P. Warren and R. L. Vande Wiele (1973), which noted 1 Chinese and 1 black person out of 42 patients seen at their New York clinic between 1960 and 1971. Other articles from the late 1970s and early 1980s mentioned one or two cases of nonwhite anorectics but did not offer any explanations of this phenomenon (Jones et al. 1980; Hedblom, Hubbard, and Andersen 1981; Garfinkel and Garner 1982; Roy-Byrne, Lee-Benner, and Yager 1984).

More recently, research on nonwhite and non-Western anorectics has grown significantly: Investigators have identified cases of the disorder in Malaysia (Buhrich 1981), Greece (Fichter, Elton, and Sourdi

1988), Nigeria (Nwaefuna 1981), Zimbabwe (Buchan and Gregory 1984), and Ethiopia (Fahy et al. 1988). The non-Western country to receive the most attention from researchers has been Japan, probably because it is one of the most Westernized East Asian countries. In Japan, anorexia nervosa and a binge-eating syndrome called *Kirbarashi-gui* have been well documented by researchers for a number of years (Nogami and Yabana 1977; Azuma and Henmi 1982; Nogami et al. 1984; Suematsu et al. 1985).

Within the United States and Great Britain, there has been a growing body of research on the incidence of anorexia nervosa in blacks (Pumariega, Edwards, and Mitchell 1984; Andersen and Hay 1985; Nevo 1985; Robinson and Andersen 1985; White, Hudson, and Campbell 1985; Silber 1986; Gray et al. 1987; Hsu 1987; Thomas and James 1988), Hispanics (Silber 1986; Hiebert et al. 1988; Smith and Krejci 1991), Asian-Americans and British-Asians (Nevo 1985; Root 1990), and Native Americans (Rosen et al. 1988; Whitehouse and Mumford 1988; Smith and Krejci 1991), as well as recent immigrants from Eastern Europe (Bulik 1987), the Middle East (Garfinkel and Garner 1982), and the Caribbean (Thomas and Szmukler 1985; Holden and Robinson 1988). Researchers have also challenged the notion that lesbians are "protected" from eating disorders because lesbian ideology challenges culturally prescribed beauty ideals (Striegel-Moore et al. 1990; Thompson 1994).

Although all of this recent research indicates that the number of reported cases of anorexia nervosa and other eating disorders is substantially lower in non-whites, lesbians, and individuals from non-Western countries, there is a difference of opinion on what this implies about the actual incidence of eating disorders in these groups. Some researchers who have investigated anorexia nervosa in racial minorities have suggested that the disorder is linked more to socioeconomic class than to race and argue that the growing incidence of nonwhite anorectics reflects the growing economic prosperity of certain minority group members. These researchers argue that as non-whites become more prosperous, their exposure to white, middle-class beauty standards increases, thereby making nonwhites more vulnerable to anorexia and other eating disorders. Because fewer nonwhites than whites belong to the middle and upper economic classes, fewer nonwhites become anorexic (Pumariega et al. 1984; Andersen and Hay 1985; Robinson and Andersen 1985; White et al. 1985; Gray et al. 1987; Hsu 1987; Thomas and James 1988).

Other investigators, however, have exposed methodological and philosophical flaws behind this kind of argument. Some have suggested that the reason there are so relatively few nonwhite anorectics is because people of color do not have the same access to health-care facilities as whites. Because most of the studies of anorexia nervosa record only those sufferers who come to the attention of medical and psychiatric

facilities, nonwhites who lack access to these facilities will not be acknowledged by health-care researchers. Moreover, even those minorities who do have access to medical care may feel threatened by a white-dominated medical profession and/or may be embarrassed to seek help for a mental health problem. Thus, the actual number of nonwhites with anorexia nervosa in the general population may be greater than indicated by case reports (Rosen et al. 1988; Root 1990; Dolan 1991; Smith and Krejci 1991; Thompson 1994).

In addition, some have argued that racial stereotypes about who is most vulnerable to anorexia nervosa can explain the apparent rarity of the disorder in minority groups. Anorexia nervosa is frequently referred to in medical and popular literature as a "Golden Girl's Disease" that afflicts only young girls from white, Western European, privileged backgrounds (Root 1990). Consequently, this ethnocentric bias may lead medical personnel to misdiagnose or underdiagnose eating disorders in persons of color (Silber 1986; Hiebert et al. 1988; Rosen et al. 1988; Dolan 1991; Thompson 1994).

Even those who agree that minority group status may "protect" nonwhites from eating disorders also argue that this status does not necessarily protect specific individuals within these groups. As Maria Root (1990: 534) notes in a recent article on eating disorders in women of color: "Individuals within each racial/ethnic group are subject to the standards of the dominant culture, particularly when the culture-of-origin is devalued by the dominant culture." Because thinness in Western and Westernized societies is associated with higher social class, and the attendant social power, resources, and opportunities, some individuals of color may see the pursuit of a slim body-type as a ticket to upward social mobility and acceptance by the dominant culture (Root 1990; see also Silber 1986 and Thompson 1994).

Whatever the explanation, the standard image of anorexia nervosa as a privileged white girl's disease is increasingly being called into question. The disorder has been detected in a variety of racial, ethnic, and socioeconomic groups and in both Western and non-Western societies, although at the moment the number of cases among these groups appears to be relatively rare. Clearly, more research is needed before any definitive statements on the incidence and form of eating disorders in nonwhite and non-Western groups can be made.

Heather Munro Prescott

Bibliography

- Anderson, Arnold, and Andy Hay. 1985. Racial and socioeconomic influences in anorexia nervosa and bulimia. *International Journal of Eating Disorders* 4: 479–87.
- Azuma, Y., and M. Henmi. 1982. A study of the incidence of anorexia nervosa in schoolgirls. *Annual report of research group in eating disorders*: 30–4.
- Bell, Rudolph. 1985. *Holy anorexia*. Chicago.
- Bliss, Eugene L., and C. H. Hardin Branch. 1960. *Anorexia nervosa: Its history, psychology, and biology*. New York.
- Bordo, Susan. 1993. *Unbearable weight: Feminism, western culture, and the body*. Berkeley, Calif.
- Brown, L. 1987. Lesbians, weight, and eating: New analyses and perspectives. In *Lesbian psychologies*, ed. Boston Lesbian Psychologies Collective, 294–310. Chicago.
- Bruch, Hilde. 1966. Anorexia nervosa and its differential diagnosis. *Journal of Nervous and Mental Disease* 141: 555–66.
1973. *Eating disorders: Obesity, anorexia nervosa, and the person within*. New York.
1978. *The golden cage: The enigma of anorexia nervosa*. New York.
1988. *Conversations with anorexics*, ed. Danita Czyzewski and Melanie A. Suhr. New York.
- Brumberg, Joan Jacobs. 1988. *Fasting girls: The emergence of anorexia nervosa as a modern disease*. Cambridge, Mass.
- Buchan, T., and L. D. Gregory. 1984. Anorexia nervosa in a black Zimbabwean. *British Journal of Psychiatry* 145: 326–30.
- Buhrich, N. 1981. Frequency and presentation of anorexia in Malaysia. *Australian and New Zealand Journal of Psychiatry* 15: 153–5.
- Bulik, Cynthia. 1987. Eating disorders in immigrants. *International Journal of Eating Disorders* 6: 133–41.
- Bynum, Caroline Walker. 1987. *Holy feast and holy fast: The religious significance of food and medieval women*. Berkeley, Calif.
- Chernin, Kim. 1985. *The hungry self: Women, eating, and identity*. New York.
- Chiple, William S. 1859. Sitomania: Its causes and treatment. *American Journal of Insanity* 26: 1–42.
- Crisp, A. H., R. L. Palmer, and R. S. Kalucy. 1976. How common is anorexia nervosa? A prevalence study. *British Journal of Psychiatry* 128: 549–54.
- Dolan, Bridget. 1991. Cross-cultural aspects of anorexia nervosa and bulimia. *International Journal of Eating Disorders* 10: 67–78.
- Fahy, T. A., P. H. Robinson, G. F. M. Russell, and B. Sheinman. 1988. Anorexia nervosa following torture in a young African woman. *British Journal of Psychiatry* 153: 385–7.
- Fenwick, S. 1880. *On atrophy of the stomach and on the nervous affections of the digestive organs*. London.
- Fichter, M. M., M. Elton, and L. Sourdi. 1988. Anorexia nervosa in Greek and Turkish adolescents. *European Archives of Psychiatric and Neurological Science* 237: 200–8.
- Freud, Sigmund. [1918] 1959. From the history of an infantile neurosis. In *Collected papers*, Vol. 3. Trans. Alix and James Strachey. New York.
- Furnham, A., and N. Alibhai. 1983. Cross-cultural differences in the perception of female body shapes. *Psychological Medicine* 13: 829–37.
- Garfinkel, Paul E., and David M. Garner. 1980. Sociocultural factors in the development of anorexia nervosa. *Psychosomatic Medicine* 10: 647–56.
1982. *Anorexia nervosa: A multidimensional perspective*. New York.
- Garner, David M., Marion Olmstead, and J. Polivy. 1983. Development and validation of a multi-dimensional eating disorder inventory of anorexia and bulimia. *International Journal of Eating Disorders* 2: 15–34.

- Gowers, Simon, and John McMahon. 1989. Social class and prognosis in anorexia nervosa. *International Journal of Eating Disorders* 8: 105-9.
- Gray, James, Kathryn Ford, and Lily Kelly. 1987. The prevalence of bulimia in a black college population. *International Journal of Eating Disorders* 6: 733-40.
- Gull, William. 1874. Anorexia nervosa (apepsia hysterica, anorexia hysterica). *Transactions of the Clinical Society of London* 7: 22-8.
- Hedblom, J., F. A. Hubbard, and A. E. Andersen. 1981. A multidisciplinary treatment program for patient and family. *Journal of Social Work in Health Care* 7: 67-86.
- Herzog, David B., and Paul M. Copeland. 1985. Eating disorders. *New England Journal of Medicine* 313: 295-303.
- Hiebert, K. A., M. A. Felice, D. L. Wingard, et al. 1988. Comparison of outcome in Hispanic and Caucasian patients with anorexia nervosa. *International Journal of Eating Disorders* 7: 693-6.
- Holden, N. L., and P. H. Robinson. 1988. Anorexia nervosa and bulimia nervosa in British blacks. *British Journal of Psychiatry* 152: 544-9.
- Hsu, George. 1987. Are eating disorders becoming more prevalent among blacks? *International Journal of Eating Disorders* 6: 113-24.
- Janet, Pierre. 1903. *Les obsessions et la psychasthénie*. New York.
- Jones, D. J., M. M. Fox, H. M. Babigian, and H. E. Hutton. 1980. Epidemiology of anorexia nervosa in Monroe County, New York, 1960-1976. *Psychosomatic Medicine* 42: 551-8.
- Kaye, W. H., H. Gwirtsman, T. George, et al. 1986. Caloric consumption and activity levels after weight recovery in anorexia nervosa: A prolonged delay in normalization. *International Journal of Eating Disorders* 5: 489-502.
- Kendall, R. E., D. J. Hall, A. Hailey, and H. M. Babigian. 1973. The epidemiology of anorexia nervosa. *Psychological Medicine* 3: 200-3.
- Lacey, J. H., and B. M. Dolan. 1988. Bulimia in British black and Asians: A catchment area study. *British Journal of Psychiatry* 152: 73-7.
- Lasègue, Charles. 1873. De l'anorexie hystérique. *Archives générales de médecine* 1: 385-7.
- Lawlor, B. A. 1987. An atypical eating disorder in a black man. *Journal of the National Medical Association* 79: 984-6.
- Lawlor, B. A., R. C. Burket, and J. A. Hodgins. 1987. Eating disorders in American black men. *Journal of the National Medical Association* 79: 984-6.
- Lawlor, B. A., and C. S. W. Rand. 1985. Bulimia nervosa in a black woman. *American Journal of Psychiatry* 142: 1514.
- Minuchin, Salvador, Bernice L. Rosman, and Lesley Baker. 1978. *Psychosomatic families: Anorexia nervosa in context*. Cambridge, Mass.
- Morton, Richard. 1689. *Phthisiologia, seu exercitationes de phthisi*. London.
- Nasser, M. 1986. Comparative study of the prevalence of abnormal eating attitudes among Arab female students of both London and Cairo Universities. *Psychological Medicine* 16: 621-5.
- Nevo, Shoshana. 1985. Bulimic symptoms: Prevalence and ethnic differences among college women. *International Journal of Eating Disorders* 4: 151-68.
- Nogami, Y., and F. Yabana. 1977. On Kibaraishi-gui. *The Japanese Journal of Psychiatry and Neurology* 31: 159-66.
- Nogami, Y., T. Yamaguchi, H. Ishiwata, et al. 1984. The prevalence of binge eating among the Japanese university population. *Journal of Nihon University Medical Association* 43: 749-52.
- Nwaefuna, A. 1981. Anorexia nervosa in a developing country. *British Journal of Psychiatry* 138: 270.
- Orbach, Susie. 1986. *Hunger strike: The anorectic's struggle as a metaphor for our age*. New York.
- Prince, Raymond. 1985. The concept of culture-bound syndrome: Anorexia nervosa and brain-fag. *Social Science and Medicine* 21: 197-203.
- Pumariega, Andres, Palmer Edwards, and Carol Mitchell. 1984. Anorexia nervosa in black adolescents. *American Academy of Child Psychiatry* 23: 111-14.
- Pyle, R. L., J. E. Mitchell, and E. D. Echart. 1981. Bulimia: A report of 34 cases. *Journal of Clinical Psychiatry* 42: 60-4.
- Robinson, Paul, and Arnold Andersen. 1985. Anorexia nervosa in American blacks. *Journal of Psychiatric Research* 19: 183-8.
- Root, Maria P. 1990. Disordered eating in women of color. *Sex Roles* 22: 525-36.
- Rosen, Lionel, Christine Shafer, Gail Dummer, et al. 1988. Prevalence of pathogenic weight-control behaviors among Native American women and girls. *International Journal of Eating Disorders* 7: 807-11.
- Rothenberg, Albert. 1986. Eating disorder as a modern obsessive-compulsive syndrome. *Psychiatry* 49: 45-53.
- Roy-Byrne, P., K. Lee-Brenner, and J. Yager. 1984. Group therapy for bulimia. *International Journal of Eating Disorders* 3: 97-116.
- Schwartz, Donald M., Michael G. Thompson, and Craig L. Johnson. 1982. Anorexia nervosa and bulimia: The socio-cultural context. *International Journal of Eating Disorders* 1: 20-36.
- Silber, Tomas. 1986. Anorexia nervosa in blacks and Hispanics. *International Journal of Eating Disorders* 5: 121-8.
- Silberstein, L. R., M. Mishkind, R. H. Striegel-Moore, et al. 1988. Behavioral and psychological implications of body dissatisfaction: Do men and women differ? *Sex Roles* 19: 219-32.
1989. Men and their bodies: A comparison of homosexual and heterosexual men. *Psychosomatic Medicine* 51: 337-46.
- Silverman, Joseph. 1977. Anorexia nervosa: Clinical and metabolic observations in a successful treatment plan. In *Anorexia nervosa*, ed. Robert A. Vigersky, 331-9. New York.
1983. Richard Morton, 1637-1698: Limner of anorexia nervosa: His life and times. *Journal of the American Medical Association* 250: 2830-2.
- Smith, Jane, and Jonathon Krejci. 1991. Minorities join the majority: Eating disturbances among Hispanic and Native American youth. *International Journal of Eating Disorders* 10: 179-86.
- Stein, A. 1989. All dressed up, but no place to go? Style wars and the new lesbianism. *Outlook* 2: 34-42.
- Striegel-Moore, Ruth, Lisa R. Silberstein, and Judith Rodin. 1986. Toward an understanding of risk factors in bulimia. *American Psychologist* 41: 246-63.
- Striegel-Moore, Ruth, Naomi Tucker, and Jeanette Hsu. 1990. Body image dissatisfaction and disordered eating. *International Journal of Eating Disorders* 9: 493-500.
- Suematsu, H., H. Ishikawa, T. Kuboki, and T. Ito. 1985. Statistical studies on anorexia nervosa in Japan: Detailed clinical data on 1,011 patients. *Psychotherapy and Psychosomatics* 43: 96-103.
- Szmukler, George I., and Digby Tantum. 1984. Anorexia nervosa: Starvation dependence. *British Journal of Medical Psychology* 57: 303-10.
- Thomas, J. P., and G. S. Szmukler. 1985. Anorexia nervosa in

- patients of Afro-Caribbean extraction. *British Journal of Psychiatry* 146: 653–6.
- Thomas, Veronica, and Michelle James. 1988. Body image, dieting tendencies, and sex role traits in urban black women. *Sex Roles* 18: 523–9.
- Thompson, Becky W. 1994. *A hunger so wide and so deep: American women speak out on eating problems*. Minneapolis, Minn.
- Warren, M. P., and R. L. Vande Wiele. 1973. Clinical and metabolic features of anorexia nervosa. *American Journal of Obstetrics and Gynecology* 117: 435–49.
- White, William C., Lisa Hudson, and Stephen J. Campbell. 1985. Bulimarexia and black women: A brief report. *Psychotherapy* 22: 449–50.
- Whitehouse, A. M., and D. B. Mumford. 1988. Increased prevalence of bulimia nervosa amongst Asian school-girls. *British Medical Journal* 297: 718.
- Willi, J., and S. Grossmann. 1983. Epidemiology of anorexia nervosa in a defined region of Switzerland. *American Journal of Psychiatry* 140: 564–7.
- Yager, J., F. Kurtzman, J. Landsverk, and E. Weismeier. 1988. Behaviors and attitudes related to eating disorders in homosexual male college students. *American Journal of Psychiatry* 145: 495–7.

IV.E.2 🐾 Celiac Disease

Historical Perspective

Celiac disease has been recognized for centuries (Dowd and Walker-Smith 1974) by physicians aware of its major symptoms of diarrhea and gastrointestinal distress accompanied by a wasting away in adults and a failure to grow in children. The Greek physician Aretaeus (first century A.D.) called the condition *coeliac diathesis* – coeliac deriving from the Greek word *koeliakos*, or abdominal cavity. The British physician Samuel Gee provided what is generally considered the first modern, detailed description of the condition, which he termed the *coeliac affection* in deference to Aretaeus, in a lecture presented at St. Bartholomew's Hospital in London (Gee 1888). At present, celiac disease (or, especially in Britain, coeliac disease) is the most commonly used term for the condition, although various others may be encountered, including celiac syndrome, celiac sprue, nontropical sprue, and gluten-sensitive enteropathy.

There were perceptions, certainly since Gee's time, that celiac disease was a consequence of, or at least affected by, diet. Gee (1888: 20) noted that “[a] child, who was fed upon a quart of the best Dutch mussels daily, thrived wonderfully, but relapsed when the season for mussels was over.” Such associations with diet led to wide-ranging dietary prescriptions and proscriptions (Haas 1924; Sheldon 1955; Weijers, Van de Kamer, and Dicke 1957; Anderson 1992). Some physicians recommended exclusion of fats – others, exclusion of complex carbohydrates. At times, so many restrictions were applied simultaneously that it became impossible to maintain a satisfactory intake of calories.

Because dietary treatments of celiac disease were of limited effectiveness, the food connection remained a puzzle until the end of the 1940s when a Dutch physician, W. K. Dicke, observed that removal of wheat from the diet of celiac patients led to dramatic improvement (Dicke 1950). It eventually became clear that the efficacy of the various diets recommended prior to Dicke's discovery was in proportion to the extent that wheat was excluded from them. Initial fractionation studies pointed to the gliadin protein fraction as being most harmful to celiac patients (Van de Kamer, Weijers, and Dicke 1953).

Soon after Dicke (1950) reported the harmful effects of wheat on celiac patients, a series of investigations indicated that rye, barley, and oats were also harmful, whereas rice and maize (corn) were not (Dicke, Weijers, and Van de Kamer 1953; Van de Kamer et al. 1953; Weijers et al. 1957). With children, and most adults, the exclusion of wheat, rye, barley, and oats from the diet usually brought about a complete, or largely complete, recovery.

During the 1950s, the development of intestinal biopsy techniques (Shiner 1956; Crosby and Kugler 1957; Brandborg, Rubin, and Quinton 1959) enabled pieces of tissue to be recovered from the intestine for examination and testing, and it was recognized that ingestion of wheat and related grains often resulted in damage to the intestinal mucosa, including the absorptive cells, or enterocytes, lining the interior surface of the intestine. The enterocytes are responsible for the absorption of almost all nutrients; the damage to them provided a basis for the gastrointestinal symptoms and malabsorption.

Familial associations, as well as its rarity among the Chinese, the Japanese, and blacks in sub-Saharan Africa (McNeish et al. 1974), indicated a likely genetic basis for celiac disease. Subsequent findings of strong associations with particular histocompatibility antigens supported this possibility (Kagnoff 1992) and, along with the presence of circulating antibodies to wheat gliadin proteins in patients on a gluten-containing diet, suggested that an abnormal immune response initiated intestinal damage in susceptible individuals.

A detailed understanding of the basis for celiac disease remains to be achieved. Neither the initiating event triggered by wheat gliadin proteins or products derived from them by digestion, nor the mechanisms leading to tissue damage in the small intestine following ingestion of wheat are completely understood. On the basis of current knowledge, it appears that celiac disease has resulted from the convergence of various evolutionary developments: These include human evolution (especially evolution of the immune system), the evolution of wheat and related grasses, the evolution of the protein structures apparently unique to wheat and closely related species, and the evolution of culture – specifically the development of agriculture and the spread of wheat farming (Figure IV.E.2.1).

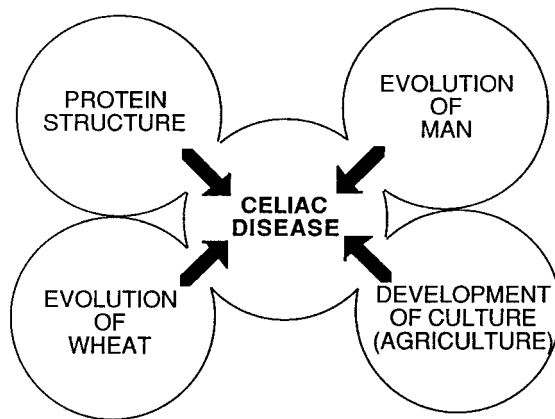


Figure IV.E.2.1. Evolutionary factors combine to produce celiac disease.

Geographical Distribution of Celiac Disease

Celiac disease is commonly thought of as largely afflicting people of European ancestry, and reports of cases among the Chinese, the Japanese (Kihara, Kukida, and Ichikawa 1977), and black Africans are sufficiently rare as to make it unlikely that this assumption is incorrect (also see McNeish et al. 1974; Simoons 1981). However, studies throughout much of the world either are inadequate or have yet to be done.

There are some parts of Asia where rice, a harmless grain (see ahead for definitions of harmful grains), is the predominant cereal grain in the diet, and in some parts of Africa, teff and millet – likely to be harmless cereal grains – predominate. To what extent a low intake of wheat, rye, barley, and oats in Asia and Africa contributes to the apparently low incidence of celiac disease is unknown. Furthermore, in some parts of the world – the United States, for example – medical personnel may have only minimal knowledge of celiac disease and fail to recognize it, whereas in Britain and much of Western Europe, physicians are much more attuned to its signs, which may be highly varied. The likelihood of failure to recognize celiac disease is doubtless substantially greater in some Asian and African countries and may contribute to the perception that it is rare in these places. The difficulties in diagnosing celiac disease and in developing statistical information about its incidence and prevalence have been discussed by R. F.A. Logan (1992a, 1992b).

The incidence of celiac disease varies throughout Europe and with time. For example, crude incidence rates (number of cases per 1,000 births) in various parts of Sweden currently range from about 2.2 to 3.5 per 1,000, whereas in neighboring Denmark, the rate is about 0.1 per 1,000 (Greco et al. 1992). Furthermore, the incidence varies with time, possibly indicat-

ing the contribution of some environmental factor. Prior to 1983, incidence of the disease in Sweden was much lower, but it began to increase at about that time, and it has been speculated that an increase in the amount of gluten in infant formulas was responsible for the rise (Maki et al. 1992). The suggestion has also been made (on the basis of screening of blood donors for antigliadin antibodies) that the incidence of celiac disease in Denmark may be similar to that of Sweden although largely undiagnosed (E. Grodzinsky, personal communication, cited by Ascher and Kristiansson 1994).

Ireland was reported to have a high incidence of celiac disease – about 4 per 1,000 births during an 11-year period prior to 1973, particularly in the west near Galway (Mylotte et al. 1973). Since that time, however, the incidence among children under the age of 12 has fallen by 62 percent (Stevens, Egan-Mitchell, et al. 1988), in contrast to the rise experienced in Sweden. The reasons for the decline in Ireland are unclear; they perhaps involve dietary changes such as an increase in the frequency of breast feeding or a decrease in the early introduction of grains to the diet of infants. Or some other undefined factor, perhaps a change in the type of viral infections prevalent in the population, may be involved.

The incidence in the United States has not been established, but a recent study of a Minnesota population (Talley et al. 1994) estimated the incidence as 0.01 per 1,000 person-years and the prevalence at 1 per 5,000. Because of the diversity of the U.S. population, which includes significant numbers of people having African, Chinese, or Southeast Asian ancestry, it might be expected that prevalence would be less than that of Europe as a whole, which is about 1 per 1,000. Whether the prevalence among U.S. residents of European extraction approximates that of European populations remains to be established, but it would be surprising to find that it is different.

Most studies of incidence and prevalence have focused on clearly recognizable disease resulting from significant damage to the absorptive epithelium of the intestine, often diagnosed by intestinal biopsy to show loss of villous structure and followed by removal of wheat and other harmful grains from the diet to demonstrate recovery of mucosal structure and function. It is becoming obvious, however, that there are many people who have subclinical celiac disease (Marsh 1992a), so that the true incidence of celiac disease may be at least two to three times greater than that represented by those with significant damage to the epithelium.

Origins of Agriculture and Wheat Farming

A diploid wheat and barley were likely to have been among the first crops cultivated by humans who were spearheading the Neolithic Revolution about 10,000 years ago. Wild diploid species of wheat, prob-

ably *Triticum boeoticum* and *Triticum urartu*, or possibly the wild tetraploid species *Triticum dicoccoides*, had most likely been harvested by hunter-gatherers for some considerable time. These species of *Triticum* were probably extant for perhaps 16 million years (Shewry et al. 1980) before domestication occurred (Harlan 1977) and thus those who pioneered in cultivating it probably had a reasonably good understanding of the plant's life cycle. The origin of wheat cultivation is likely to have been somewhere in the Fertile Crescent area of the Middle East, perhaps near Jericho.

Frederick Simoons (1981) has discussed the possible effects of the spread of wheat farming (from likely centers of origin in the Middle East) on populations with members genetically susceptible to celiac disease. He used a genetic marker for celiac disease and compared the incidence of this marker in populations throughout the world with the incidence of celiac disease. The marker selected was a human leukocyte antigen, HLA-B8 (a class I major histocompatibility complex [MHC] protein). HLA-B8 had been demonstrated to have approximately threefold greater occurrence in celiac patients (Falchuk, Rogentine, and Strober 1972; Stokes et al. 1972) than in normal individuals of the same populations, although it is now known that certain class II MHC proteins show even better correlations and are more likely to play a direct role in the processes of celiac disease (Kagnoff, Morzycka-Wroblewska, and Harwood 1994). The data available to Simoons were, however, limited to HLA-B8.

Simoons reasoned that the greater rate of occurrence of celiac disease among people with the HLA-B8 antigen gave researchers a tool to use in their study of the illness. By determining the distribution of the antigen among human populations, researchers could predict which people were more at risk of developing the disease. Although there are exceptions, such that populations without HLA-B8 may yet have high levels of celiac disease (Logan 1992a), the assumption appears to have general validity.

It turned out that there were two geographic centers in Old World populations with noticeably higher frequencies of HLA-B8, one in western Europe, centered about the British Isles, and one in northwestern India, centered about the Punjab region. These regions also fell within the major areas of wheat cultivation, and in both there had been reports of relatively high incidences of celiac disease. They also represented, to a considerable degree, the boundaries of the spread of wheat farming throughout its geographical range up to relatively recent times (in this analysis, about 1000 B.C.) from its point of origin.

There appeared to be a gradient in the incidence of HLA-B8 throughout the range of wheat cultivation, such that low levels of the antigen were found at the likely origin of wheat cultivation and higher levels at the periphery. Simoons (1981) assumed that high levels of HLA-B8 were once typical of peoples at the ori-

gin and suggested that the gradient observed from the center to the periphery might well reflect the effects of a selective genetic disadvantage to members of the population carrying HLA-B8 – along with other genes for susceptibility to celiac disease. Thus, the high levels of HLA-B8 in populations that took up wheat farming late (in the west of Ireland, for example) in its spread from the center of origin, along with high incidences of celiac disease, would reflect the lesser time available for the selective disadvantage to have diminished marker levels.

One obvious discrepancy was northwestern China, where wheat farming was introduced quite late through trade. This major wheat-growing area, although somewhat beyond the contiguous area of the earlier spread of wheat farming, might reasonably be considered its periphery.

There was, however, neither evidence of celiac disease nor high levels of HLA-B8 in the wheat-growing regions of China. A possible explanation for this was the likelihood that immune systems evolved differently in different populations to protect against relatively specific infectious diseases. Geographically distant populations would have been exposed to different stresses, thereby resulting in different complements of histocompatibility antigens. Furthermore, histocompatibility antigens that provided resistance to one disease might, quite coincidentally, enhance susceptibility to another, and this may well be true of genes for susceptibility to celiac disease (Strober 1992). Thus, the absence of celiac disease in the wheat-growing areas of China may well reflect the absence of the susceptibility genes in the Chinese population. Although hampered by a lack of adequate information, the analysis by Simoons (1981) is at least a highly interesting attempt to deal with what must have been selective pressures on populations containing the genes for susceptibility to celiac disease as wheat farming spread from the Middle East throughout Western Europe.

Definition of Harmful Cereal Grains and Components

Wheat, Rye, Barley, and Oats

Early conclusions regarding the toxicity of cereal grains and their constituents were based mainly on the ability of a grain to produce malabsorption of fats in celiac patients. Despite the lack of the more sophisticated approaches available today, early test results generally seem convincing. Currently, examination of mucosal biopsy specimens has become fairly common. When a flattened mucosa is found, with loss of villous structure, wheat and other harmful grains are removed from the diet to see if improvement follows. If it does, a subsequent challenge and biopsy to test for relapse may follow, although subsequent challenge has come to be reserved for special circumstances (Walker-Smith et al. 1990) because supporting antibody tests (tests for cir-

culating antigliadin and antiendomysium antibodies) have lessened the likelihood of misdiagnosis (McMillan et al. 1991). In recent years, however, the situation has become more complicated through recognition that a flattened mucosa may actually be an extreme response and that there are many gluten-sensitive people who show less obvious evidence of the disease (Marsh 1992a). Circulating antigliadin antibodies may be indicative of celiac disease or at least a related gluten-sensitive condition in the absence of any evidence of mucosal changes, particularly when symptoms are present (O'Farrelly 1994).

Relatively few subjects were used in the testing of some grains or grain fractions in early work. This has been a continuing problem, resulting largely from the difficulties inherent in the requirement for human subjects in celiac disease research. It has become fairly clear that response to challenge may vary greatly from one patient to another and for a single patient over time. Furthermore, a considerably delayed response to challenge is not unusual, even when the challenge is with wheat, presumably the most toxic of the cereal grains (Egan-Mitchell, Fottrell, and McNicholl 1978; Walker-Smith, Kilby, and France 1978).

The variations in response known to occur might explain the opposing conclusions arrived at by various investigators regarding the harmful effects of oats. For example, Dicke, H. A. Weijers, and J. H. Van de Kamer (1953) asserted that oats are toxic, whereas W. Sheldon (1955) concluded that they are not. Part of the problem apparently arises from the relatively small proportion of avenins, the likely toxic fraction, in oat grain proteins (Peterson and Brinegar 1986): Avenins may make up only about 10 percent of the total protein in oats. Furthermore, when small numbers of subjects are involved, the average response of one group can be quite different from that of another. Because a negative response in one or a few patients might indicate only that the feeding time was too short, a clear positive response, as in the studies of oat protein toxicity by P. G. Baker and A. E. Read (1976), should perhaps be given more weight than a negative response, as in the study by A. S. Dissanayake, S. C. Truelove, and R. Whitehead (1974). The latter effort, however, included testing by intestinal biopsy – the more rigorous test – whereas the former did not.

A recent feeding trial using oats with 10 confirmed celiac patients produced no harmful effects as indicated by changes in mucosal architecture, endomysial antibody development, and infiltration of intraepithelial lymphocytes (Srinivasan et al. 1996). The relatively low percentage of avenins in oats may complicate this study in that a daily dosage of 50 grams (g) of oats was fed for 3 months. This would correspond to about 5 g of avenins per day. If 50 g of wheat had been fed instead, the patients would be eating at least 40 g of gliadins per day. Nevertheless, subsequent challenge of two of the patients with only 0.5 g of gluten per day produced evidence of intestinal dam-

age. Thus, the toxicity of oats must be considered questionable on the basis of this latest study, as there are now two careful studies that showed no evidence of toxicity and only one study that was positive. It should also be emphasized that the supposedly positive study did not include biopsies.

In other cases, the test materials may not have been well defined. It is possible that the maize (corn) flour used in early testing was actually maize starch and relatively free of protein (Simoons 1981); the conclusion that maize proteins are not harmful may not have as rigorous a scientific base as is generally thought. There is, however, no obvious reason to question the apparent safety of maize.

A lack of sensitivity of the available testing methods, combined with inadequate test length, may have been responsible for the early conclusions that wheat starch is safe for celiac patients (Dicke et al. 1953). These conclusions, however, were challenged in later work (Ciclitira, Ellis, and Flagg 1984; Freedman et al. 1987a, 1987b; Skerrett and Hill 1990, 1991) as a consequence of the development of sensitive monoclonal antibody tests for gliadins (but also see Booth et al. 1991). Such tests have demonstrated the presence of small amounts of gliadins in wheat starch preparations, although the question of how harmful these small amounts may be to celiac patients is controversial (Ejderhamn, Veress, and Strandvik 1988; Hekkens and van Twist-de Graaf 1990; Campbell 1992).

It may be that rigorous scientific studies have proved toxicity in celiac disease only for wheat (and wheat proteins). To a considerable extent, conclusions regarding the toxicity of rye and barley – and, conversely, the lack of toxicity for rice and maize – are not based on an adequate amount of rigorous scientific testing. (Rigorous studies of oats are beginning to appear, and these support an absence of toxicity for this grain.)

Wheat Evolution

Wheat is a member of the grass family (Gramineae), as are the other grains – rye, barley, and oats – that are suspected of harming people with celiac disease. Of course, triticale, a cross between wheat and rye, is toxic, as would be expected for any other similar crosses that included genetic material from one or more of the toxic grains.

The grasses are a relatively recent evolutionary development. They are angiosperms, flowering plants, that developed somewhere between 100 and 200 million years ago (Cronquist 1968). Fossil evidence for grasses goes back only about 65 million years. They became widespread during the Oligocene epoch about 25 to 40 million years ago (Thomasson 1980), which, when it is considered that life has been evolving on earth for about 4 billion years, is a relatively short time. Cereal grains presumably evolved more recently within the time frame of grass evolution.

The ancestors of wheat, barley, and rye were

diploid species with 7 chromosomes in the haploid state, or 14 chromosomes in vegetative cells. Barley first diverged from the common ancestral line, followed by rye. The line eventually gave rise to a series of diploid species that may be classed as *Triticum* species (Morris and Sears 1967). A natural cross occurred at some unknown time between two slightly diverged *Triticum* species, presumably *Triticum urartu* and *Triticum speltooides*, that gave rise to a new species through a process of polyploid formation in which chromosomes become doubled.

Without chromosome doubling, a cross between related species, such as those of *Triticum*, can take place, but the offspring is infertile because of a failure by the chromosomes to pair during meiosis. Polyploid formation occurs naturally, albeit rarely, perhaps through a process involving unreduced gametes (Harlan and de Wet 1975). This process can also be achieved in the laboratory through the use of chemicals (such as colchicine) that interfere with spindle formation to block chromosomal separation at anaphase. Either way, the result is a doubled set of chromosomes, a condition that allows for each original chromosome set to pair with its identical replicated set, thus avoiding the chromosome pairing problems that occur in crosses between species. The result of allopolyploid formation between *T. urartu* and *T. speltooides* was a new fertile species, called a tetraploid because it incorporated four sets of chromosomes in vegetative cells (designated AABB). Durum wheats (*Triticum turgidum* var. *durum*), used for pasta making, belong to one of the varietal groups of the AABB tetraploids.

Some time after wheat cultivation began, a tetraploid wheat crossed accidentally with a weed growing in the same field (*Triticum tauschii* var. *strangulata*), followed by chromosome doubling to give rise to a hexaploid species, which, because the genome of *T. tauschii* was designated D, had the composition AABBDD. Bread wheats (*Triticum aestivum* var. *aestivum*) produce a more elastic dough than durum wheats and have properties that lend themselves to the retention of gases produced during yeast fermentation (leavening) of doughs. It was in the context of the evolution of wheat and closely related species that proteins arose with amino acid sequences capable of initiating damage in persons with celiac disease.

Gluten Protein Evolution

The gluten proteins of wheat (a monocotyledonous plant) and certain closely related species – including rye, barley, and, with qualifications, oats – are unique among plant storage proteins in having exceptionally high proportions of the amino acids glutamine and proline. It was suggested (Kasarda 1980, 1981) that because of their unique composition, the extensive occurrence of repeating sequences based largely on glutamine and proline, and their occurrence only in

recently evolved grass species, wheat prolamins are a late evolutionary development. Subsequently, with the development of molecular biological techniques, gene sequencing provided evidence of short amino acid sequences in most wheat prolamins, homologous to sequences found in storage globulins of more distantly related dicotyledonous plant species (Kreis et al. 1985). This discovery pushes the possible age of the ancestral genes back to within the period during which the flowering plants have existed – perhaps 100 million years.

However, repetitive sequences make up major parts of all gluten proteins, whereas the homologous sequences just mentioned occurred only in nonrepetitive regions. These repetitive sequences, having major proportions of glutamine and proline, apparently do not have counterparts in proteins of species outside the grass family. Accordingly, it seems likely that at least the repetitive domains of gluten proteins, which have slightly differing repeat motifs according to type, are of more recent (<~65 million years) origin. All of the various repeating sequences include glutamine and proline residues, and although the repeats are often imperfect, comparison allows a consensus repeating sequence to be recognized. It is at least possible that the most active (in celiac disease) peptides may result from variations on the themes represented by the consensus sequences, as a consequence of these imperfections, rather than from the consensus sequences themselves, but that remains to be established.

There is reasonably strong evidence that the peptides with sequences found in the repeat domain of alpha-gliadin are capable of triggering the intestinal damage characteristic of celiac disease. Furthermore, significant sequence similarities between gluten proteins and other proteins are rare. The first important similarity found was between alpha-gliadin and the E1b protein (Kagnoff et al. 1984), produced in conjunction with infection by adenovirus 12, which infects the human gastrointestinal tract. Recently, similarities have been found for peptides produced in conjunction with infection by other types of adenovirus (Lähdeaho et al. 1993).

The recent evolution of repeating sequence domains in gluten proteins through extensive duplication of the DNA codons (for glutamine and proline, along with a few other amino acids) corresponding to the repeat motifs may be the basis for the lack of homologies or similarities with other proteins. Most proteins do not have large amounts of glutamine and proline. Hence, the sequences active in celiac disease are likely to be confined to the grass family.

Fractionation of Wheat

Gluten

The wheat kernel is made up largely of the endosperm, the interior part of the grain, which constitutes about 85 percent of the kernel. In the milling

process, the crushed kernels are fractionated by sieving. The crushed endosperm is the source of white flour, and the outer layers of the kernel yield the bran and germ fractions. Endosperm cells are largely made up of starch granules (about 75 percent) surrounded by a proteinaceous matrix. The proteins of this surrounding matrix are mainly storage proteins; upon germination of the seed, they are broken down to provide a source of nitrogen for use by the new plant. The storage proteins are to a large extent the same as the proteins of gluten, which contribute elasticity, yet at the same time extensibility, to flour-water doughs.

Gluten (Beccari 1745), prepared by washing away the starch granules from dough, constitutes the resulting cohesive, elastic mass of protein. It usually retains about 15 percent albumins and globulins, which are water- and salt-soluble proteins, respectively, according to the classification system of T. B. Osborne (1907). The remainder of the protein may be considered to constitute true gluten proteins, notable for their high percentages of two amino acids, glutamine and proline, which provided the basis for the name prolamins (Osborne 1907). Glutamine usually makes up 30 to 55 percent, and proline 15 to 30 percent, of the amino acids in gluten proteins. These two amino acids seem fairly certain to be included in the amino acid sequences of gluten proteins that are active in celiac disease.

The Meaning of "Gluten-Free"

Celiac patients, the physicians who treat them, and the various organizations that represent celiac patients usually indicate that a food safe for someone with celiac disease is "gluten-free." This use of the term "gluten" is sometimes confusing. Traditionally, gluten, as just mentioned, is the cohesive, elastic protein obtained when a wheat flour dough is kneaded in excess water to wash away the starch granules (Beccari 1745). The resultant gluten ball can only be obtained readily from wheat-flour doughs, with difficulty from rye, and probably not at all from barley or oats. Cereal chemists would not ordinarily refer to "rye gluten" or "barley gluten," let alone "oat gluten," but would refer specifically to the equivalent storage proteins in these grains (for example, the proteins equivalent to the gliadin proteins of wheat are called hordeins in barley, secalins in rye, and avenins in oats). The term gluten has become corrupted in recent times, however, occasionally being used in industry to refer to the protein residues from other grains. For example, when maize is separated into starch and protein fractions, the protein fraction is often referred to as "corn gluten."

What celiac patients wish to know about a food when they ask if it contains gluten is: Does it contain wheat, rye, barley, or oat proteins or any of the harmful peptides that are derived from the storage proteins of these grains during food processing or through the action of the digestive enzymes? Use of

the term gluten-free is particularly awkward when applied to seeds of rather distantly related plants, such as the dicots amaranth and quinoa. These plants are not known to have proteins at all similar to gluten proteins. However, because quite small peptides of specific amino acid sequence may be active in celiac disease, it is not beyond possibility that equivalent sequences might occur in generally quite different proteins of distantly related species. Despite its ambiguities, the term gluten-free in relation to foods for celiac patients is already well established, and finding a satisfactory and short alternative may be difficult.

Digestion and Nutrient Absorption

Digestion of proteins commences in the stomach with partial denaturation (unfolding) of the large protein polypeptide chains by hydrochloric acid, which enhances their breakdown into smaller polypeptide chains by the proteolytic enzyme pepsin. In the highly acidic conditions of the stomach, pepsin is very active, cleaving a number of different peptide bonds. Breakdown of food proteins in the stomach is incomplete (Gray 1991), and relatively large peptides pass from the stomach into the duodenal part of the small intestine. Enzymes secreted by the pancreas – trypsin and carboxypeptidases – enter the interior (lumen) of the small intestine to continue digestion of proteins until they are broken down into amino acids or small peptides. Peptides this small are unlikely to be active in celiac disease.

Breakdown of proteins to amino acids or short peptides, both of which can be absorbed, is probably complete in the distal small bowel, even in celiac patients, as evidenced by a gradient of intestinal damage from the duodenum through the jejunum in active celiac disease. The ileum may remain free of tissue damage (Rubin et al. 1962). Instillation of wheat proteins into the ileum (Rubin et al. 1962) or the rectum (Dobbins and Rubin 1964; Loft, Marsh, and Crowe 1990) produces similar damage to that seen in the small intestine, indicating that the entire epithelial surface is susceptible to the damaging effects of wheat proteins or peptides. The final breakdown of food particles to relatively small molecules, and their active or passive absorption for use by the body, takes place at the membranes of cells lining the surface of the small intestine.

Almost all nutrients, ranging through amino acids, sugars, fatty acids, vitamins, and minerals are absorbed from the small intestine so that damage to this absorptive surface may have many manifestations in celiac disease. Obviously, a deficiency of calories from carbohydrates and fats and a deficiency of amino acids needed for protein synthesis can be responsible for a loss of weight or a failure to thrive, but the effects of malabsorption are often more diverse, ranging from osteoporosis in later life as a consequence of a failure to absorb calcium adequately to nerve degen-

eration as a consequence of a failure to absorb vitamins. The large intestine, or colon, extracts water and electrolytes from the food residues while bacterial action on these residues reduces the bulk of indigestible materials, such as cellulose. Relatively little absorption of nutrients occurs in the colon.

When intestinal biopsy was introduced in the 1950s, it was demonstrated that the surface of the intestinal mucosa in patients with active celiac disease appeared flattened as a consequence of the loss of villous structure and enhanced proliferation of immature cells in the crypts. Subsequently it was recognized that the microvilli of mature enterocytes were often damaged as well. Together, these losses significantly diminish the absorptive surface area and give rise to an increased crypt layer with an immature population of enterocytes having incompletely developed enzyme and transport activities. The net result is a diminished capability to absorb nutrients, although it should be noted that malabsorption may occur even in patients with little or no obvious damage to the epithelium.

It has been noted that initial responses to gluten challenge included infiltration by lymphocytes of the epithelium and lamina propria and thickening of the basement membrane and that these changes preceded major changes in mucosal structure (Shmerling and Shiner 1970; Marsh 1992a), providing strong support for the involvement of immunological processes in the destruction of the absorptive epithelium. A highly significant association of celiac disease with certain proteins of the major histocompatibility complex (histocompatibility antigens) also provides evidence for involvement of the immune system and a genetic predisposition to the disease. The inheritance of celiac disease appears to be complex, however, involving two or more genes (Strober 1992), and an environmental contribution, such as a viral infection or stress, may be necessary before the genetic predisposition comes into play (Kagnoff et al. 1984). What remains to be clarified is how these immune processes are triggered by gliadin peptides and how they ultimately result in the loss of epithelial absorptive cells.

Mechanisms of Tissue Damage

The way in which ingested gluten protein triggers events in the body that may ultimately bring about damage to the absorptive epithelium of the small intestine is unknown. The first important hypothesis put forward postulated that a key enzyme, a protease that could degrade proteins, was absent in celiac patients. As a consequence, certain harmful gliadin peptides that would, in the normal person, have been broken down and rendered harmless, continued to exist with a consequence of either direct toxicity on the absorptive epithelial cells or initiation of immune responses that secondarily damaged the epithelium

(Cornell 1988). Although it is difficult to disprove this hypothesis beyond any reasonable doubt, no missing enzyme has ever been found despite considerable research effort to locate it. Furthermore, the likelihood that limited digestion of proteins in the stomach results in exposure of at least the proximal small intestine (duodenum) to fairly large peptides raises the question of why most people do not suffer some damage from such peptides.

The missing enzyme hypothesis was largely supplanted by the proposal that binding of gluten proteins or peptides to the enterocytes targeted them specifically for destruction and that enhanced proliferation of crypt cells, in response to destruction of the villous enterocytes, resulted in flattening of the mucosal surface, one of the major characteristics of advanced disease. Unfortunately, no evidence for the binding of gluten proteins or peptides to enterocytes *in vivo* has been found. The finding that enterocytes were capable of expressing MHC class II antigens that might be capable of presenting small gluten peptides of perhaps 10 to 20 amino acid residues to T cells raises the question of how effective the search for the binding of peptides to enterocytes has been, insofar as the methods used (often based on antibody probes produced in response to stimulation with intact proteins) were unlikely to recognize such small gluten peptides.

The strong correlation of celiac disease with particular class II histocompatibility antigens, which present peptides to T cells (lymphocytes), thereby activating them, has resulted in a currently favored hypothesis in which this presentation to T cells and binding of gliadin/gluten peptides to T cells is the initiating process that results in damage to the mucosal tissues, especially those underlying the epithelial cells (enterocytes). Binding of peptides by T cells may, however, activate a number of pathways, and beyond the activation step, the details of the hypothesis become rather vague.

Genes and Proteins of the Major Histocompatibility Complex

It is clear that there is an increased incidence of celiac disease in certain families, providing support for the hypothesis that there is a genetic predisposition to the disease. We continue to stress, however, that this predisposition may be insufficient for development of the disease without the intervention of some environmental factor – other than gluten proteins.

Celiac disease occurs in about 14 percent of siblings, 8 percent of parents, and 15 percent of children of celiac patients (Strober 1992). Although dizygotic (fraternal) twins do not show major differences from normal siblings in their tendency to develop celiac disease, monozygotic (identical) twins are about 70 percent concordant for the disease. The fact that 30 percent of the identical twins are discordant, with

one having the disease and the other being free of it, is strong support for the necessary role of an environmental factor, such as viral infection (Kagnoff et al. 1984) or some other stressful event.

It was recognized in conjunction with attempts to transplant tissues from one individual to another, that certain protein antigens (human leukocyte antigens, or HLA molecules) had to be matched to avoid rejection of the foreign tissue. These proteins, now often called the proteins of the major histocompatibility complex (MHC proteins, MHC antigens), are coded by a cluster of genes on human chromosome 6. MHC proteins are cell-surface receptors, proteins that bind peptides at specific binding sites for presentation to a receptor site on T cells. The T-cell receptor must interact simultaneously with both the MHC molecule and the peptide presented by the MHC molecule before the presentation is recognized as a signal for the T cell to carry out some other function. An activated T cell might go on, for example, to activate B lymphocytes to produce antibodies, or it must suppress the normal immune response to antigens encountered as part of our food intake (development of oral tolerance). The latter function may be especially important because, presumably, in most people the immune response to gluten peptides has been suppressed, whereas in celiac patients, certain gluten peptides trigger an immune response capable of damaging the small intestine.

The MHC proteins of concern to us here are divided into class I and class II on the basis of a number of distinguishing characteristics, one of which is the type of cell on which they are expressed. Class I antigens appear on most cells, whereas class II antigens are found mainly on cells of the immune system, although class II antigens are expressed on the surfaces of enterocytes as well. The first demonstrations of associations between MHC proteins and celiac disease indicated that a class I antigen, designated HLA-B8, was found in nearly 90 percent of celiac patients but in only about 20 percent of controls (Falchuk, Rogentine, and Strober 1972). This antigen, HLA-8, was the genetic marker that, as was discussed earlier by Simoons, was thought to determine the incidence of celiac disease throughout the world.

However, subsequently it became clear that associations of celiac disease with class II antigens were even stronger than those of the class I antigen HLA-B8. Particularly strong associations have been found for the HLA class II alleles DR3 and DQ2. The association with HLA-B8 apparently resulted from linkage disequilibrium in which certain closely linked genes tend to remain associated during genetic recombinations to a greater extent than would be expected.

The associations of celiac disease with various MHC proteins observed throughout the world are quite complex (see reviews of Kagnoff 1992; Tighe and Ciclitira 1993), but they are sufficiently strong to make it seem likely that class II antigens are directly

involved in the mechanisms responsible for tissue damage in celiac disease – possibly by presenting peptides derived from gluten proteins to T cells.

In support of the possibility that gliadin peptide binding by T cells is involved in celiac disease, K. E. A. Lundin and colleagues (1993) have found that T-cell lines derived from intestinal biopsies recognized gliadin peptides, and the cell lines were mainly stimulated by antigen presentation in the context of the DQ heterodimer. In addition, H. A. Gjertson, Lundin, and colleagues (1994) found that a peripheral blood lymphocyte clone from a celiac patient specifically recognized a synthetic peptide corresponding to alpha-gliadin residues 31-47. Further interest in the role of class II MHC proteins derives from studies showing that HLA-DR molecules, which are expressed by enterocytes of normal individuals or of celiac patients on a wheat-free diet (Arnaud-Battandier et al. 1986; Ciclitira et al. 1986), became differentially expressed in the crypt cells of celiac patients (but not in those of normal controls). This occurred when gluten proteins or peptides were added to biopsied tissues in culture that had been obtained from treated patients on a wheat-free diet (Fais et al. 1992).

W. Strober (1992) has pointed out that because discordance for celiac disease is much less common in identical twins than in siblings who have apparently identical complements of histocompatibility antigens (MHC proteins), it seems likely that the difference cannot be explained by an environmental factor because the contribution of environmental effects should have been about the same for the sets of twins and siblings. Strober considers this as strong evidence for a contribution by some non-MHC-defined disease gene. Whether or not this gene turns out to lie outside the MHC complex, celiac disease is likely to have a two-locus basis. D. A. Greenberg, S. E. Hodge, and J. I. Rotter (1982) considered that the available data fit best with a recessive-recessive two-gene model, but studies in the west of Ireland (Hernandez et al. 1991) favored dominance for the gene associated with the HLA locus. Conclusions may be somewhat dependent on the population studied, as many diseases show variation in their genetic basis.

Associated Diseases

Many different diseases have been reported to occur concurrently with celiac disease, including dermatitis herpetiformis, insulin-dependent diabetes, Down syndrome, IgA nephropathy, and epilepsy associated with cerebral calcifications (Bayless, Yardley, and Hendrix 1974; Gobbi et al. 1992; Collin and Maki 1994). In only a few diseases are statistically significant data available for the establishment of an association with celiac disease.

Dermatitis herpetiformis is a disease manifested by a rash with small blisters and IgA deposits even in uninvolved skin (Fry 1992). It is quite strongly associ-

ated with celiac disease. Intestinal lesions similar to those encountered in celiac disease are found upon biopsy of about two-thirds of patients with dermatitis herpetiformis, and a gluten-free diet usually improves both the skin rash and the intestinal lesions. Furthermore, even those patients without significant damage to the intestinal epithelium usually have an increased number of intraepithelial lymphocytes, and patients without obvious mucosal damage have developed such damage when their intake of gluten was increased. Nevertheless, malabsorption is much less common in dermatitis herpetiformis than in celiac disease (Kumar 1994). Indeed, the incidence is rather less than that of celiac disease, with dermatitis herpetiformis occurring about half as frequently as celiac disease in Sweden and perhaps one-fifth as frequently in Scotland (Edinburgh).

Insulin-dependent diabetes mellitus also shows a definite association with celiac disease. Between 5 and 10 percent of children with celiac disease have diabetes mellitus, whereas about 1 to 3 percent of children with diabetes mellitus have celiac disease (Visakorpi 1969; Strober 1992). Both celiac disease and insulin-dependent diabetes mellitus share an association with the histocompatibility antigen HLA-DR3 (Maki et al. 1984), and the common MHC gene(s) may predispose individuals carrying them to both conditions.

Down syndrome may be weakly associated with celiac disease. J. Dias and J. Walker-Smith (1988) found an increased incidence of Down syndrome in celiac patients compared with the incidence in the general population. They considered this as supporting evidence for earlier findings (Nowak, Ghisham, and Schulze-Delrieu 1983). M. Castro and colleagues (1993) have confirmed a significant increase in celiac disease among Down syndrome patients and found that antigliadin antibodies provide a useful screening tool to look for celiac disease among these patients. In addition, there are reports that patients with schizophrenia, autism, and IgA nephropathy have apparently benefited by removing wheat (and related harmful grains) from their diets, although such reports are not universally accepted as valid.

There have been extensive attempts to show a correlation between schizophrenia and celiac disease (Dohan 1966, 1988; Lorenz 1990), but the results have not been convincing. Evidence for intestinal damage in the group of schizophrenics studied by F. M. Stevens and colleagues (1977) was no greater than that for controls, and there was no significant increase in serum antireticulin antibodies. Both intestinal damage and antireticulin antibodies are usually present in celiac patients taking a normal (wheat-containing) diet. Nevertheless, published reports from physicians indicating that removal of wheat from the diet can produce a marked reduction in psychotic symptoms for some schizophrenics were considered sufficiently convincing by K. Lorenz (1990) to indicate that in at

least a subset of schizophrenics, wheat has an adverse effect on the disease. Even in studies where no positive correlation was found, investigators noted that some patients did apparently show considerable improvement on a wheat-free ("gluten-free") diet (Rice, Ham, and Gore 1978).

Some celiac patients on a gluten-free diet, perhaps about 25 percent (Ansaldi et al. 1988), have indicated that they feel temporary psychological disturbance upon eating wheat, including symptoms such as irritability, hostility, depression, or a general feeling of mental unease. Although such symptoms may be "normal" in the face of physiological changes produced by eating wheat, investigation of a connection between such responses in celiac patients and the effects of wheat in the diet of carefully selected schizophrenics might be worthwhile. Again, even if wheat in the diet does adversely affect the course of schizophrenia, the mechanisms involved may be quite different from those involved in producing intestinal damage in celiac disease.

The role of wheat in autism is also controversial, but development of the hypothesis that wheat and casein exacerbate the symptoms of autistic patients parallels that for schizophrenia. The hypothesis is based so far on studies that indicate value for a wheat-free diet in improving behavior of patients with autism (Reichelt et al. 1991), although these studies are not well controlled because of the difficulties in carrying them out with patients who may not be able to supply informed consent. It has been discovered that amino acid sequences found in the primary structures of wheat gliadin proteins and casein proteins are similar to those of endorphins and other neuroactive peptides termed exorphins (Zioudrou, Streaty, and Klee 1979). It has been proposed that these, along with other neuroactive peptides from wheat, are responsible for the exacerbation of symptoms claimed for schizophrenic patients on a normal, wheat-containing diet (Zioudrou et al. 1979; Dohan 1988). This working hypothesis is also favored by some researchers investigating the possibility that wheat and casein proteins in the diet exacerbate symptoms in autistic patients (Reichelt et al. 1991). According to the hypothesis, food-protein-derived neuroactive peptides pass through the wall of the intestine and also pass the blood-brain barrier to affect brain function. This results in a variety of abnormal behaviors.

Abnormal peptide patterns appear in the urine of autistic subjects and schizophrenics (Cade et al. 1990; Shattock et al. 1990; Reichelt et al. 1991), and the hypothesis has been put forward that these abnormal patterns reflect to some extent the abnormal absorption and then excretion of exorphin peptides derived from wheat or milk proteins in the diet. This implies either excessive passage through the intestinal wall and/or a failure in some other way to process these peptides into harmless forms. However, that these

abnormal urine peptide patterns truly represent excretion of peptides of dietary origin does not seem to have been proved.

IgA nephropathy is a kidney disease characterized by protein and blood in the urine and IgA deposits in the kidney (Stevens, Lavelle, et al. 1988). Although the severity is variable, the disease can lead to chronic renal failure. Reports of concurrence with celiac disease, the presence of circulating IgA antibodies to gluten proteins, and patient improvement (even with nonceliac patients) on a gluten-free diet have led to speculation that there may be a connection between the two diseases (Coppo et al. 1986; Sategna-Guidetti et al. 1992). It has been suggested that IgA nephropathy might be similar to dermatitis herpetiformis, with the main difference being that IgA deposits form in the kidney rather than in the skin. However, the evidence indicates that epithelial damage is uncommon in IgA nephropathy, although the activities of marker enzymes in the brush border were significantly lower (Stevens, Lavelle, 1988). Gliadin proteins or peptides have been found in complex with the IgA deposits in IgA nephropathy (Russel et al. 1986), and similar associations should be sought in dermatitis herpetiformis.

Celiac patients have been found to be at increased risk of developing certain types of cancer (Holmes et al. 1989; Logan et al. 1989; Holmes and Thompson 1992), especially small intestinal lymphoma. Nonetheless, the absolute risk of a celiac patient dying from this cancer is small because the incidence in the normal population is quite low.

G. K. T. Holmes and colleagues (1989) found in a study of 210 patients that those who had been on a strict gluten-free diet for more than 5 years did not have significantly increased risk of developing lymphoma. But the risk of developing lymphoma was reported to be high (1 in 10) for celiac patients who were diagnosed late in life (over the age of 50). The investigators did not discuss what constituted a "strict gluten-free diet," and J. A. Campbell (1992) has pointed out that it would not have been unusual for patients who thought themselves to be on a strict gluten-free diet to be using products containing wheat starch, which has a small amount of gluten in it. It seems prudent for celiac patients to follow a strict gluten-free diet as recommended by Holmes and colleagues (1989), but whether traces of gluten in the diet, such as might result from use of wheat-starch products, contribute to the development of malignancies late in life does not appear to have been established.

Conclusion

Human evolution has produced a complex immune system designed to protect us from harmful parasites, bacteria, viruses, and foreign substances. Human agricultural development has led to a dependence on agriculture, and in a large part of the world, a heavy dependence on wheat and related grain crops as part

of that agriculture. The proteins of these grain crops have evolved in a somewhat isolated manner (in part because they are recently evolved) such that particular amino acid sequences have appeared in these proteins that can induce the immune system in a susceptible subset of people into damaging the absorptive layer of cells lining the small intestine. The mechanisms are not very well understood, but they seem to involve a failure of the suppression mechanisms by which oral tolerance to food-protein antigens is developed in normal individuals.

Donald D. Kasarda

Part of this chapter is excerpted from the author's "Toxic Cereal Grains in Coeliac Disease" (1994), pp. 203–20, in *Gastrointestinal Immunology and Gluten-Sensitive Disease*, ed. C. Feighery and C. O'Farrelly. The author thanks Professor Conleth Feighery, of Trinity College in the University of Dublin, Ireland, for permission to refer to unpublished results achieved by him and his colleagues, and for his comments on the manuscript.

Bibliography

- Altschul, S. F., W. Gish, W. Miller, et al. 1990. Basic local alignment search tool. *Journal of Molecular Biology* 215: 403–10.
- Ammerman, A. J., and L. L. Cavalli-Sforza. 1971. Measuring the rate of spread of early farming in Europe. *Man* 6: 674–88.
- Anderson, C. M. 1992. The evolution of a successful treatment for coeliac disease. In *Coeliac disease*, ed. M. N. Marsh, 1–16. London.
- Anderson, C. M., A. C. Frazer, J. M. French, et al. 1952. Coeliac disease: Gastrointestinal studies and the effect of dietary wheat flour. *Lancet* 1: 836–42.
- Ansaldi, N., K. Tavassoli, D. Faussone, et al. 1988. Clinical and histological behaviour of previously diagnosed children with coeliac disease when rechallenged with gluten. In *Coeliac disease: 100 years*, ed. P. J. Kumar and J. Walker-Smith, 237–9. Leeds, England.
- Arnaud-Battandier, N. Cerf-Bensussan, R. Amsellem, and J. Schmitz. 1986. Increased HLA-DR expression by enterocytes in children with celiac disease. *Gastroenterology* 91: 1206–12.
- Ascher, H., and B. Kristiansson. 1994. Childhood coeliac disease in Sweden. *Lancet* 344: 340–1.
- Auricchio, S., G. De Ritis, M. De Vincenzi, et al. 1984. Agglutinating activity of gliadin-derived peptides from bread wheat: Implications for coeliac disease pathogenesis. *Biochemical and Biophysical Research Communications* 121: 428–33.
- Baker, P. G., and A. E. Read. 1976. Oats and barley toxicity in coeliac patients. *Postgraduate Medical Journal* 52: 264–8.
- Bayless, T. M., J. H. Yardley, and T. R. Hendrix. 1974. Coeliac disease and possible disease relationships. In *Coeliac disease: Proceedings of the Second International Coeliac Symposium*, ed. W. Th. J. M. Hekkens and A. S. Peña, 351–9. Leiden, the Netherlands.
- Beccari, J. B. 1745. De frumento. *De Bononiensi Scientiarum et Artium Instituto atque Academia Commentarii*, 7 vols. II, Part 1, 122–7. Bologna.
- Bernardin, J. E., R. M. Saunders, and D. D. Kasarda. 1976.

- Absence of carbohydrate in celiac-toxic A-gliadin. *Cereal Chemistry* 53: 612-14.
- Bietz, J. A. 1982. Cereal prolamins evolution revealed by sequence analysis. *Biochemical Genetics* 20: 1039-53.
- Blanco, A., P. Resta, R. Simeone, et al. 1991. Chromosomal location of seed storage protein genes in the genome of *Dasypyrum villosum* (L.) Candargy. *Theoretical and Applied Genetics* 82: 358-62.
- Booth, C. C., M. S. Losowsky, J. A. Walker-Smith, and J. D. W. Whitney. 1991. Inter-laboratory variation in gluten detection by ELISA kit. *Lancet* 337: 1094.
- Brandborg, L. L., C. E. Rubin, and W. E. Quinton. 1959. A multipurpose instrument for suction biopsy of the esophagus, stomach, small bowel, and colon. *Gastroenterology* 37: 1-16.
- Bronstein, H. D., L. J. Haeflner, and O. D. Kowlessar. 1966. Enzymatic digestion of gliadin: The effect of the resultant peptides in adult celiac disease. *Clinica Chimica Acta* 14: 141-55.
- Bruce, G., J. F. Woodley, and C. J. H. Swan. 1984. Breakdown of gliadin peptides by intestinal brush borders from coeliac patients. *Gut* 25: 919-24.
- Cade, R., H. Wagemaker, R. M. Privette, et al. 1990. The effect of dialysis and diet on schizophrenia. In *Psychiatry: A World Perspective*, Vol. 3, ed. C. N. Stefanis, A. D. Rabavilas, and C. R. Soldatos, 494-500. New York.
- Campbell, J. A. 1992. Dietary management of celiac disease: Variations in the gluten-free diet. *Journal of the Canadian Dietetic Association* 53: 15-18.
- Castro, M., A. Crino, B. Papadatou, et al. 1993. Down's syndrome and celiac disease: The prevalence of high IgA-antigliadin antibodies and HLA-DR and DQ antigens in trisomy 21. *Journal of Pediatric Gastroenterology and Nutrition* 16: 265-8.
- Chicz, R. M., R. G. Urban, W. S. Lane, et al. 1992. Predominant naturally processed peptides bound to HLA-DR1 are derived from MHC-related molecules and are heterogeneous in size. *Nature* 358: 764-8.
- Ciclitira, P. J., H. J. Ellis, and N. L. K. Fagg. 1984. Evaluation of a gluten free product containing wheat gliadin in patients with coeliac disease. *British Medical Journal* 289: 83.
- Ciclitira, P. J., D. J. Evans, N. L. K. Fagg, et al. 1984. Clinical testing of gliadin fractions in coeliac disease. *Clinical Science* 66: 357-64.
- Ciclitira, P. J., J. M. Nelufer, H. J. Ellis, and D. J. Evans. 1986. The effect of gluten on HLA-DR in the small intestinal epithelium of patients with coeliac disease. *Clinical and Experimental Immunology* 63: 101-4.
- Collin, P., and M. Maki. 1994. Associated disorders in coeliac disease: Clinical aspects. *Scandinavian Journal of Gastroenterology* 29: 769-75.
- Coppo, R., B. Bassolo, C. Rollino, et al. 1986. Dietary gluten and primary IgA nephropathy. *New England Journal of Medicine* 315: 1167-8.
- Cornell, H. J. 1974. Gliadin degradation and fractionation. In *Coeliac disease: Proceedings of the Second International Coeliac Symposium*, ed. W. Th. J. M. Hekkens and A. S. Peña, 74-5. Leiden, the Netherlands.
1988. Wheat proteins and celiac disease. *Comments in Agricultural and Food Chemistry* 1: 289-313.
- Cronquist, A. 1968. *The evolution of flowering plants*. Boston.
- Crosby, W. H., and H. W. Kugler. 1957. Intraluminal biopsy of the small intestine: The intestinal biopsy capsule. *American Journal of Digestive Diseases* 2: 236-41.
- Cuadro, C., G. Ayet, C. Burbano, et al. 1995. Occurrence of saponins and sapogenols in Andean crops. *Journal of the Science of Food and Agriculture* 67: 169-72.
- De Ritis, G., S. Auricchio, H. W. Jones, et al. 1988. In vitro (organ culture) studies of the toxicity of specific A-gliadin peptides in celiac disease. *Gastroenterology* 94: 41-9.
- De Ritis, G., P. Occorsio, S. Auricchio, et al. 1979. Toxicity of wheat flour proteins and protein-derived peptides for in vitro developing intestine from rat fetus. *Pediatric Research* 13: 1255-61.
- Dias, J., and J. Walker-Smith. 1988. Down's syndrome and coeliac disease. In *Coeliac disease: 100 years*, ed. P. J. Kumar and J. A. Walker-Smith, 277-9. Leeds, England.
- Dicke, W. K. 1950. Coeliakie: Een onderzoek naar de nadelige invloed van sommige graansoorten op de lijder aan coeliakie. M.D. thesis, University of Utrecht, Utrecht, the Netherlands.
- Dicke, W. K., H. A. Weijers, and J. H. Van de Kamer. 1953. Coeliac disease. II. The presence in wheat of a factor having a deleterious effect in cases of celiac disease. *Acta Paediatrica* (Stockholm) 42: 34-42.
- Dissanayake, A. S., S. C. Truelove, and R. Whitehead. 1974. Lack of harmful effects of oats on small-intestinal mucosa in coeliac disease. *British Medical Journal* 4: 189-91.
- Dobbins, W. O., and C. E. Rubin. 1964. Studies of the rectal mucosa in celiac sprue. *Gastroenterology* 47: 471-9.
- Dohan, F. C. 1966. Cereals and schizophrenia: Data and hypothesis. *Acta Psychiatrica Scandinavica* 42: 125-32.
1988. Genetic hypothesis of idiopathic schizophrenia: Its exorphin connection. *Schizophrenia Bulletin* 14: 489-94.
- Dowd, B., and J. Walker-Smith. 1974. Samuel Gee, Aretaeus, and the coeliac affection. *British Medical Journal* 2: 45-7.
- Dvorak, J., D. D. Kasarda, M. D. Dietler, et al. 1986. Chromosomal location of seed storage protein genes in the genome of *Elytrigia elongata*. *Canadian Journal of Genetics and Cytology* 28: 818-30.
- Egan-Mitchell, B., P. F. Fottrell, and B. McNicholl. 1978. Prolonged gluten tolerance in treated coeliac disease. In *Perspectives in coeliac disease*, ed. B. McNicholl, C. F. McCarthy, and P. F. Fottrell, 251-7. Baltimore, Md.
- Ejderhamn, J., B. Veress, and B. Strandvik. 1988. The long-term effect of continual ingestion of wheat starch-containing gluten-free products in coeliac patients. In *Coeliac disease: 100 years*, ed. P. J. Kumar and J. Walker-Smith, 294-7. Leeds, England.
- Engler, J. A., M. S. Hoppe, and M. P. van Bree. 1983. The nucleotide sequence of the genes encoded in early region 2b of human adenovirus type 7. *Gene* 21: 145-59.
- Fais, S., L. Maiuri, F. Pallone, et al. 1992. Gliadin induces changes in the expression of MHC-class II antigens by human small intestinal epithelium. Organ culture studies with coeliac mucosa. *Gut* 33: 472-5.
- Falchuk, G. M., R. L. Gebhard, C. Sessoms, and W. Strober. 1974. An in vitro model of gluten-sensitive enteropathy. *Journal of Clinical Investigation* 53: 487-500.
- Falchuk, Z. M., D. L. Nelson, A. J. Katz, et al. 1980. Gluten-sensitive enteropathy: Influence of histocompatibility type on gluten sensitivity in vitro. *Journal of Clinical Investigation* 66: 227-33.
- Falchuk, Z. M., F. N. Rogentine, and W. Strober. 1972. Predominance of histocompatibility antigen HL-A8 in patients with gluten-sensitive enteropathy. *Journal of Clinical Investigation* 51: 1602-5.
- Fluge, Ø., K. Sletten, G. Fluge, et al. 1994. In vitro toxicity of purified gluten peptides tested by organ culture. *Journal of Pediatric Gastroenterology and Nutrition* 18: 186-92.
- Forssell, F., and H. Wieser. 1994. Dinkelweizen und Zöliakie. *Bericht 1994, Deutsche Forschungsanstalt für Lebensmittelchemie*. Garching, Germany.

- Frazer, A. C. 1956. Discussion on some problems of steatorrhea and reduced stature. *Proceedings of the Royal Society of Medicine* 49: 1009-13.
- Frazer, A. C., M. B. Fletcher, C. A. C. Ross, et al. 1959. Gluten-induced enteropathy: The effect of partially-digested gluten. *Lancet* 2: 252-5.
- Freedman, A. R., G. Galfre, E. Gal, et al. 1987a. Detection of wheat gliadin contamination of gluten-free foods by a monoclonal antibody dot immunobinding assay. *Clinica Chimica Acta* 166: 323-8.
- 1987b. Monoclonal antibody ELISA to quantitate wheat gliadin contamination of gluten-free foods. *Journal of Immunological Methods* 98: 123-7.
- Fry, L. 1992. Dermatitis herpetiformis. In *Coeliac disease*, ed. M. N. Marsh, 81-104. London.
- Gautam, A. M., C. B. Lock, D. E. Smilek, et al. 1994. Minimum structural requirements for peptide presentation by major histocompatibility complex class II molecules: Implications in induction of immunity. *Proceedings of the National Academy of Sciences (U.S.)* 91: 767-71.
- Gee, Samuel. 1888. On the coeliac affection. *St. Bartholomew's Hospital Reports* 24: 17-20.
- Gjertsen, H. A., K. E. A. Lundin, L. M. Sollid, et al. 1994. T cells recognize a peptide derived from alpha-gliadin presented by the celiac disease associated HLA-DQ (alpha 1*0501, beta 1*0201) heterodimer. *Human Immunology* 39: 243-52.
- Gjertsen, H. A., L. M. Sollid, J. Ek, et al. 1994. T cells from the peripheral blood of coeliac disease patients recognize gluten antigens when presented by HLA-DR, -DQ, or -DP molecules. *Scandinavian Journal of Immunology* 39: 567-74.
- Gobbi, G., F. Bouquet, L. Greco, et al. 1992. Coeliac disease, epilepsy, and cerebral calcifications. *Lancet* 340: 439-43.
- Gray, G. 1991. Dietary protein processing: Intraluminal and enterocyte surface effects. In *Handbook of physiology: The gastrointestinal system IV*, 411-20. New York.
1992. Starch digestion and absorption in nonruminants. *Journal of Nutrition* 122: 172-7.
- Greco, L., M. Maki, F. Di Donato, and J. K. Visakorpi. 1992. Epidemiology of coeliac disease in Europe and the Mediterranean area. A summary report on the multicentre study by the European Society of Pediatric Gastroenterology and Nutrition. In *Common food intolerances 1: Epidemiology of coeliac disease*, ed. S. Auricchio and J. K. Visakorpi, 25-44. Basel.
- Greenberg, D. A., S. E. Hodge, and J. I. Rotter. 1982. Evidence for recessive and against dominant inheritance at the HLA-"linked" locus in celiac disease. *American Journal of Human Genetics* 34: 263-77.
- Haas, S. V. 1924. Value of banana treatment of celiac disease. *American Journal of Diseases of Children* 28: 420-1.
- Harlan, J. R. 1977. The origins of cereal agriculture in the Old World. In *Origins of agriculture*, ed. C. A. Reed, 357-83. The Hague.
- Harlan, J. R., and J. M. J. de Wet. 1975. On Ö. Winge and a prayer: The origins of polyploidy. *Botanical Review* 41: 361-90.
- Hekkens, W. Th. J. M. 1978. The toxicity of gliadin: A review. In *Perspectives in coeliac disease*, ed. B. McNicholl, C. F. McCarthy, and P. E. Fottrell, 3-14. Baltimore, Md.
- Hekkens, W. Th. J. M., A. J. Ch. Haex, and R. G. J. Willighagen. 1970. Some aspects of gliadin fractionation and testing by a histochemical method. In *Coeliac disease*, ed. C. C. Booth and R. H. Dowling, 11-19. London.
- Hekkens, W. Th. J. M., C. J. Van den Aarsen, J. P. Gilliams, Ph. Lems-Van Kan, and G. Bouma-Frölich. 1974. Alpha-gliadin structure and degradation. In *Coeliac disease: Proceedings of the Second International Coeliac Symposium*, ed. W. Th. J. M. Hekkens and A. S. Peña, 39-45. Leiden, Netherlands.
- Hekkens, W. Th. J. M., and M. van Twist-de Graaf. 1990. What is gluten-free - levels and tolerances in the gluten-free diet. *Die Nabrung* 34: 483-7.
- Hernandez, J. L., J. P. Michalski, C. C. McCombs, et al. 1991. Evidence for a dominant gene mechanism underlying celiac disease in the west of Ireland. *Genetic Epidemiology* 8: 13-28.
- Hitchcock, A. S. 1950. *Manual of the grasses of the United States*. Second edition, rev. A. Chase. USDA Miscellaneous Publication No. 200. Washington, D.C.
- Holmes, G. K. T., P. Prior, M. R. Lane, et al. 1989. Malignancy in coeliac disease - effect of a gluten free diet. *Gut* 30: 333-8.
- Holmes, G. K. T., and H. Thompson. 1992. Malignancy as a complication of coeliac disease. In *Coeliac disease*, ed. M. N. Marsh, 105-35. London.
- Jos, J., L. Charbonnier, J. Mossé, et al. 1982. The toxic fraction of gliadin digests in coeliac disease. Isolation by chromatography on Biogel P-10. *Clinica Chimica Acta* 119: 263-74.
- Jos, J., L. Charbonnier, J. F. Mougnot, et al. 1978. Isolation and characterization of the toxic fraction of wheat gliadin in coeliac disease. In *Perspectives in coeliac disease*, ed. B. McNicholl, C. F. McCarthy, and P. E. Fottrell, 75-89. Baltimore, Md.
- Jos, J., M. F. de Tand, F. Arnaud-Battandier, et al. 1983. Separation of pure toxic peptides from a beta-gliadin subfraction using high-performance liquid chromatography. *Clinica Chimica Acta* 134: 189-98.
- Kagnoff, M. F. 1992. Genetic basis of coeliac disease: Role of HLA genes. In *Coeliac disease*, ed. M. N. Marsh, 215-38. London.
- Kagnoff, M. F., R. F. Austin, J. J. Hubert, et al. 1984. Possible role for a human adenovirus in the pathogenesis of celiac disease. *Journal of Experimental Medicine* 160: 1544-7.
- Kagnoff, M. F., J. I. Harwood, T. L. Bugawan, and H. A. Erlich. 1989. Structural analysis of the HLA-DR, -DQ, and -DP alleles on the celiac disease-associated HLA-DR3 (DRw17) haplotype. *Proceedings of the National Academy of Sciences (U.S.)* 86: 6274-7.
- Kagnoff, M. F., E. Morzycka-Wroblewska, and J. I. Harwood. 1994. Genetic susceptibility to coeliac disease. In *Gastrointestinal immunology and gluten-sensitive disease*, ed. C. Feighery and C. O'Farrelly, 131-44. Dublin.
- Karagiannis, J. A., J. D. Priddle, and D. P. Jewell. 1987. Cell-mediated immunity to a synthetic gliadin peptide resembling a sequence from adenovirus 12. *Lancet* 1: 884-6.
- Kasarda, D. D. 1980. Structure and properties of alpha-gliadins. *Annales de Technologie Agricole (Paris)* 29: 151-73.
1981. Toxic proteins and peptides in celiac disease: Relations to cereal genetics. In *Food, nutrition, and evolution: Food as an environmental factor in the genesis of human variability*, ed. D. W. Walcher and N. Kretchmer, 201-16. New York.
- Kasarda, D. D., T. W. Okita, J. E. Bernardin, et al. 1984. Nucleic acid (cDNA) and amino acid sequences of alpha-type gliadins from wheat (*Triticum aestivum*). *Proceedings of the National Academy of Sciences of the United States of America* 81: 4712-16.
- Katz, S. I., Z. M. Falchuk, M. V. Dahl, et al. 1972. HL-A8: A genetic link between dermatitis herpetiformis and gluten-sensitive enteropathy. *Journal of Clinical Investigation* 51: 2977-80.

- Kendall, M. J., P. S. Cox, R. Schneider, and C. F. Hawkins. 1972. Gluten subfractions in coeliac disease. *Lancet* 2: 1065–7.
- Kihara, T., S. Kukida, and Y. Ichikawa. 1977. Ultrastructural studies of the duodenal epithelium of Japanese celiac sprue. *Journal of Clinical Electron Microscopy* 10: 5–6.
- Krainick, H. G., and G. Mohn. 1959. Weitere Untersuchungen über den schädlichen Weizenmehleffekt bei der Coliakie. 2. Die Wirkung der enzymatischen Abbauprodukte des Gliadin. *Helvetica Paediatrica Acta* 14: 124–40.
- Kreft, I., and B. Javornik. 1979. Buckwheat as a potential source of high quality protein. In *Seed protein improvement in cereals and grain legumes*, Vol. 2, 377–83. International Atomic Energy Agency. Vienna.
- Kreis, M., P. R. Shewry, B. G. Forde, et al. 1985. Structure and evolution of seed storage proteins and their genes with particular reference to those of wheat, rye, and barley. *Oxford Surveys of Plant Molecular and Cell Biology* 2: 253–317.
- Kumar, P. 1994. Dermatitis herpetiformis. In *Gastrointestinal immunology and gluten-sensitive disease*, ed. C. Feighery and C. O'Farrelly, 158–68. Dublin.
- Kumar, P. J., T. S. Sinclair, M. J. G. Farthing, et al. 1984. Clinical testing of pure gliadins in coeliac disease. *Gastroenterology* 86: 1147–52.
- Lafiandra, D., S. Benedettelli, B. Mariotta, and E. Porceddu. 1989. Chromosomal location of gliadin coding genes in *T. aestivum* ssp. *spelta* and evidence on the lack of components controlled by Gli-2 loci in wheat aneuploids. *Theoretical and Applied Genetics* 78: 177–83.
- Lähdeaho, M.-L., P. Parkkonen, T. Reunala, et al. 1993. Antibodies to adenovirus E1b protein derived peptides of enteric adenovirus type 40 are associated with coeliac disease and dermatitis herpetiformis. *Clinical Immunology and Immunopathology* 69: 300–5.
- Lajolo, F. M., U. M. Lanfer Marquez, T. M. C. C. Filesetti-Cozzi, and D. Ian McGregor. 1991. Chemical composition and toxic compounds in rapeseed (*Brassica napus* L.) cultivars grown in Brazil. *Journal of Agricultural and Food Chemistry* 39: 1933–7.
- Lew, E. J.-L., D. D. Kuzmicky, and D. D. Kasarda. 1992. Characterization of LMW-glutenin subunits by RP-HPLC, SDS-PAGE, and N-terminal amino acid sequencing. *Cereal Chemistry* 69: 508–15.
- Loft, D. E., M. N. Marsh, and P. Crowe. 1990. Rectal gluten challenge and the diagnosis of coeliac disease. *Lancet* 335: 1293–5.
- Logan, R. F. A. 1992a. Epidemiology of celiac disease. In *Coeliac disease*, ed. M. N. Marsh, 192–214. London.
- 1992b. Problems and pitfalls in the epidemiology of coeliac disease. In *Common food intolerances 1: Epidemiology of coeliac disease*, ed. S. Auricchio and J. K. Visakorpi, 14–24. Basel.
- Logan, R. F. A., E. A. Rifkind, I. D. Turner, and A. Ferguson. 1989. Mortality in celiac disease. *Gastroenterology* 97: 265–71.
- Lorenz, K. 1990. Cereals and schizophrenia. In *Advances in cereal science and technology*, Vol. 10, ed. Y. Pomeranz, 435–69. St. Paul, Minn.
- Lundin, K. E. A., H. Scott, T. Hansen, et al. 1993. Gliadin-specific, HLA-DQ (a1*0501, b1*0201) restricted T cells isolated from the small intestinal mucosa of celiac disease patients. *Journal of Experimental Medicine* 178: 187–96.
- Maki, M., O. Hällström, T. Huupponen, et al. 1984. Increased prevalence of coeliac disease in diabetes. *Archives of Disease in Childhood* 59: 739–42.
- Maki, M., K. Holm, H. Ascher, and L. Greco. 1992. Factors affecting clinical presentation of coeliac disease: Role of type and amount of gluten-containing cereals. In *Common food intolerances 1: Epidemiology of coeliac disease*, ed. S. Auricchio and J. K. Visakorpi, 76–82. Basel.
- Maki, M., S. Koskimies, and J. K. Visakorpi. 1988. Latent coeliac disease. In *Coeliac disease: 100 years*, ed. P. J. Kumar and J. A. Walker-Smith, 245. Leeds, England.
- Mantzaris, G., and D. P. Jewell. 1991. In vivo toxicity of a synthetic dodecapeptide from A gliadin in patients with coeliac disease. *Scandinavian Journal of Gastroenterology* 26: 392–8.
- Marsh, M. N. 1992a. Gluten, major histocompatibility complex, and the small intestine. *Gastroenterology* 102: 330–54.
- 1992b. Mucosal pathology in gluten sensitivity. In *Coeliac disease*, ed. M. N. Marsh, 136–91. London.
- McMillan, S. A., D. J. Houghton, J. D. Biggart, et al. 1991. Predictive value for coeliac disease of antibodies to gliadin, endomysium, and jejunum in patients attending for biopsy. *British Medical Journal* 303: 1163–5.
- McNeish, A. S., C. J. Rolles, R. Nelson, et al. 1974. Factors affecting the differing racial incidence of coeliac disease. In *Coeliac disease: Proceedings of the Second International Conference*, ed. W. Th. J. M. Hekkens and A. S. Peña, 330–6. Leiden, the Netherlands.
- Messer, M., C. M. Anderson, and L. Hubbard. 1964. Studies on the mechanism of destruction of the toxic action of wheat gluten in coeliac disease by crude papain. *Gut* 5: 295–303.
- Morris, R., and E. R. Sears. 1967. The cytogenetics of wheat and its relatives. In *Wheat and wheat improvement*, ed. K. S. Quisenberry and L. P. Reitz, 19–87. Madison, Wis.
- Mylotte, M., B. Egan-Mitchell, C. F. McCarthy, and B. McNicholl. 1973. Incidence of coeliac disease in the west of Ireland. *British Medical Journal* 24: 703–5.
- Nowak, T. V., F. K. Ghisham, and K. Schulze-Delrieu. 1983. Celiac sprue in Down's syndrome: Considerations on a pathogenic link. *American Journal of Gastroenterology* 78: 280–3.
- O'Farrelly, C. 1994. Wheat protein sensitive disease may be independent of intestinal structural changes. In *Gastrointestinal immunology and gluten-sensitive disease*, ed. C. Feighery and C. O'Farrelly, 169–80. Dublin.
- Offord, R. E., B. S. Anand, J. Piris, and S. C. Truelove. 1978. Further fractionation of digests of gluten. In *Perspectives in coeliac disease*, ed. B. McNicholl, C. F. McCarthy, and P. F. Fottrell, 25–9. Baltimore, Md.
- Osborne, T. B. 1907. *The proteins of the wheat kernel*. Washington, D.C.
- Peterson, D., and A. C. Brinegar. 1986. Oat storage proteins. In *Oats: Chemistry and technology*, ed. F. H. Webster, 153–203. St. Paul, Minn.
- Rahnotra, G. S., J. A. Gelroth, B. K. Glaser, and K. J. Lorenz. 1995. Baking and nutritional qualities of a spelt wheat sample. *Lebensmittel-Wissenschaft und -Technologie* 28: 118–22.
- Reichelt, K. L., A. M. Knivsberg, G. Lind, and M. Nødland. 1991. Probable etiology and possible treatment of childhood autism. *Brain Dysfunction* 4: 308–19.
- Rice, J. R., C. H. Ham, and W. E. Gore. 1978. Another look at gluten in schizophrenia. *American Journal of Psychiatry* 135: 1417–18.
- Rubenstein, I., and D. E. Geraghty. 1986. The genetic organization of zein. In *Advances in cereal science and technology VIII*, 297–315. St. Paul, Minn.
- Rubin, C. E., L. H. Brandborg, A. M. Flick, et al. 1962. Studies of celiac sprue. III. The effect of repeated wheat instillation into the proximal ileum of patients on a gluten-free diet. *Gastroenterology* 43: 621–41.

- Russel, M. W., J. Mestecky, B. A. Julian, and J. H. Galla. 1986. IgA-associated renal diseases: Antibodies to environmental antigens in sera and deposition of immunoglobulins and antigens in glomeruli. *Journal of Clinical Immunology* 6: 74-9.
- Sategna-Guidetti, C., G. Ferfaglia, M. Bruno, et al. 1992. Do IgA antigliadin and IgA antiendomysial antibodies show there is latent celiac disease in primary IgA nephropathy? *Gut* 33: 476-8.
- Shattock, P., A. Kennedy, F. Rowell, and T. Berney. 1990. Role of neuropeptides in autism and their relationships with classical neurotransmitters. *Brain Dysfunction* 3: 328-45.
- Sheldon, W. 1955. Coeliac disease. *Lancet* 2: 1097-1103.
- Shewry, P. R., J.-C. Autran, C. C. Nimmo, et al. 1980. N-terminal amino acid sequence homology of storage protein components from barley and a diploid wheat. *Nature* 286: 520-2.
- Shewry, P. R., S. Parmar, and D. J. C. Pappin. 1987. Characterization and genetic control of the prolamins of *Haynaldia villosa*: Relationship to cultivated species of the Triticeae (rye, wheat, and barley). *Biochemical Genetics* 25: 309-25.
- Shewry, P. R., P. A. Sabelli, S. Parmar, and D. Lafiandra. 1991. Alpha-type prolamins are encoded by genes on chromosomes 4Ha and 6Ha of *Haynaldia villosa* Schur (syn. *Daspyrum villosum* L.). *Biochemical Genetics* 29: 207-11.
- Shewry, P. R., S. J. Smith, E. J.-L. Lew, and D. D. Kasarda. 1986. Characterization of prolamins from meadow grasses: Homology with those of wheat, barley, and rye. *Journal of Experimental Botany* 37: 633-9.
- Shewry, P. R., A. S. Tatham, and D. D. Kasarda. 1992. Cereal proteins and celiac disease. In *Coeliac disease*, ed. M. N. Marsh, 305-48. London.
- Shiner, M. 1956. Duodenal biopsy. *Lancet* 1: 17-19.
- Shmerling, D. H., and M. Shiner. 1970. The response of the intestinal mucosa to the intraduodenal instillation of gluten in patients with coeliac disease in remission. In *Coeliac disease*, ed. C. C. Booth and R. H. Dowling, 64-74. Edinburgh.
- Shorrosh, B. S., L. Wen, K. C. Sen, et al. 1992. A novel cereal storage protein: Molecular genetics of the 19kDa globulin of rice. *Plant Molecular Biology* 18: 151-4.
- Simoons, F. J. 1981. Celiac disease as a geographic problem. In *Food, nutrition, and evolution: Food as an environmental factor in the genesis of human variability*, ed. D. N. Walcher and N. Kretschmer, 179-99. New York.
- Sjøström, H., S. U. Friis, O. Norén, and D. Anthonsen. 1992. Purification and characterization of antigenic gliadins in coeliac disease. *Clinica Chimica Acta* 207: 227-37.
- Skerritt, J. H., and A. S. Hill. 1990. Monoclonal antibody sandwich enzyme-immunoassays for determination of gluten in foods. *Journal of Agricultural and Food Chemistry* 38: 1771-8.
1991. Self-management of dietary compliance in coeliac disease by means of ELISA "home test" to detect gluten. *Lancet* 337: 379-82.
1992. How "free" is "gluten-free"? Relationship between Kjeldahl nitrogen values and gluten protein content for wheat starches. *Cereal Chemistry* 69: 110-12.
- Sollid, L., G. Markussen, J. Ek, et al. 1989. Evidence for a primary association of coeliac disease to a particular HLA-DQ a/b heterodimer. *Journal of Experimental Medicine* 169: 345-50.
- Srinivasan, U., N. Leonard, E. Jones, et al. 1996. Absence of oats toxicity in adult coeliac disease. *British Medical Journal* 313: 1300-1.
- Sterchi, E. E., and J. F. Woodley. 1977. Peptidases of the human intestinal brush border membrane. In *Perspectives in coeliac disease*, ed. B. McNicholl, C. F. McCarthy, and P. F. Fottrell, 437-49. Baltimore, Md.
- Stern, M., A. Stallmach, B. Gellerman, and H. Wieser. 1988. In vitro testing of gliadin peptides. In *Coeliac disease: 100 years*, ed. P. J. Kumar and J. A. Walker-Smith, 148-51. Leeds, England.
- Stevens, F. M., B. Egan-Mitchell, E. Cryan, et al. 1988. Decreasing incidence of coeliac disease. In *Coeliac disease: 100 years*, ed. P. J. Kumar and J. A. Walker-Smith, 306. Leeds, England.
- Stevens, F. M., E. Lavelle, M. Kearns, et al. 1988. Is IgA nephropathy (Berger's disease) dermatitis herpetiformis of the kidney? In *Coeliac disease: 100 years*, ed. P. J. Kumar and J. A. Walker-Smith, 280-3. Leeds, England.
- Stevens, F. M., R. S. Lloyd, S. M. J. Geraghty, et al. 1977. Schizophrenia and coeliac disease - the nature of the relationship. *Psychological Medicine* 7: 259-63.
- Stokes, P. L., P. Asquith, G. K. T. Holmes, et al. 1972. Histocompatibility antigens associated with adult coeliac disease. *Lancet* 1: 162-4.
- Strober, W. 1992. The genetic basis of gluten-sensitive enteropathy. In *The genetic basis of common diseases*, ed. R. A. King, J. I. Rotter, and A. Motulsky, 279-304. New York.
- Sturgess, R., P. Day, H. J. Ellis, et al. 1994. Wheat peptide challenge in coeliac disease. *Lancet* 343: 758-61.
- Talley, N. J., M. Valdovinos, T. M. Petterson, et al. 1994. Epidemiology of celiac sprue: A community-based study. *American Journal of Gastroenterology* 89: 843-6.
- Tao, H. P., and D. D. Kasarda. 1989. Two-dimensional gel mapping and N-terminal sequencing of LMW-glutenin subunits. *Journal of Experimental Botany* 40: 1015-20.
- Thomasson, J. R. 1980. Paleoagrostology: A historical review. *Iowa State Journal of Research* 54: 301-17.
- Tighe, M. R., and P. J. Ciclitira. 1993. The implications of recent advances in coeliac disease. *Acta Paediatrica* 82: 805-10.
- Van de Kamer, J. H., and H. A. Weijers. 1955. Coeliac disease. V. Some experiments on the cause of the harmful effect of wheat gliadin. *Acta Paediatrica* 44: 465-9.
- Van de Kamer, J. H., H. A. Weijers, and W. K. Dicke. 1953. Coeliac disease. IV. An investigation into the injurious constituents of wheat in connection with their action on patients with coeliac disease. *Acta Paediatrica* (Stockholm) 42: 223-31.
- Van der Hulst, R. R. W. J., B. K. van Kreel, M. F. von Meyenfelt, et al. 1993. Glutamine and the preservation of gut integrity. *Lancet* 341: 1363-5.
- Van Roon, J. H., A. J. Ch. Haex, W. A. Seeder, and J. de Jong. 1960. Chemical and biochemical analysis of gluten toxicity. No. II. *Gastroenterologia* 94: 227-35.
- Visakorpi, J. K. 1969. Diabetes and coeliac disease. *Lancet* 2: 650.
- Walker-Smith, J. A., S. Guandalini, J. Schmitz, et al. 1990. Revised criteria for diagnosis of coeliac disease. *Archives of the Diseases of Childhood* 65: 909-11.
- Walker-Smith, J. A., A. Kilby, and N. E. France. 1978. Reinvestigation of children previously diagnosed as coeliac disease. In *Perspectives in coeliac disease*, ed. B. MacNicholl, C. F. McCarthy, and P. F. Fottrell, 267-76. Baltimore, Md.
- Wang, S.-Z., and A. Esen. 1986. Primary structure of a proline-rich zein and its cDNA. *Plant Physiology* 81: 70-4.
- Weijers, H. A., J. H. Van de Kamer, and W. K. Dicke. 1957. Celiac disease. In *Advances in pediatrics*, Vol. 9, ed. S. Z. Levine, 277-318. Chicago.
- Wieser, H., and H.-D. Belitz. 1992. Isolation and fragmentation

- of the coeliac-active gliadin peptide CT-1. *Zeitschrift für Lebensmittel-Untersuchung und -Forschung* 194: 229–34.
- Wieser, H., H.-D. Belitz, and A. Ashkenazi. 1984. Amino acid sequence of the coeliac active peptide B 3142. *Zeitschrift für Lebensmittel-Untersuchung und -Forschung* 179: 371–6.
- Wieser, H., H.-D. Belitz, A. Ashkenazi, and D. Idar. 1983. Isolation of coeliac active peptide fractions from gliadin. *Zeitschrift für Lebensmittel-Untersuchung und -Forschung* 176: 85–94.
- Wieser, H., H.-D. Belitz, D. Idar, and A. Ashkenazi. 1986. Coeliac activity of the gliadin peptides CT-1 and CT-2. *Zeitschrift für Lebensmittel-Untersuchung und -Forschung* 182: 115–17.
- Wieser, H., G. Springer, H.-D. Belitz, et al. 1982. Toxicity of different wheat gliadins in coeliac disease. *Zeitschrift für Lebensmittel-Untersuchung und -Forschung* 175: 321–6.
- Woodley, J. F. 1972. Pyrrolidonecarboxyl peptidase activity in normal intestinal biopsies and those from coeliac patients. *Clinica Chimica Acta* 42: 211–13.
- Zhai, C. K., X. L. Jiang, Y. S. Xu, and K. J. Lorenz. 1994. Protein and amino acid composition of Chinese and North American wild rice. *Lebensmittel-Wissenschaft und -Technologie* 27: 380–3.
- Zioudrou, C., R. A. Sreaty, and W. A. Klee. 1979. Opioid peptides derived from food proteins: The exorphins. *The Journal of Biological Chemistry* 254: 2446–9.
- Zubillaga, P., J. C. Vitoria, A. Arrieta, et al. 1993. Down's syndrome and celiac disease. *Journal of Pediatric Gastroenterology and Nutrition* 16: 168–71.

IV.E.3 Food Allergies

Adverse Food Reactions and Allergies

An adverse food reaction is defined as any untoward reaction following the ingestion of food (Lifshitz 1988). These reactions generally fall into two categories: food intolerance and food hypersensitivity. Intolerances are nonimmunologic (Sampson and Cooke 1990) and are responsible for most adverse food reactions. They may be idiosyncratic, due to metabolic disorders, or caused by pharmacological substances such as toxins or drugs present in food. Food additives can also be a cause of food intolerance or hypersensitivity and can produce respiratory and gastrointestinal complaints (Lifshitz 1988).

True food hypersensitivity, or allergy, is an adverse food reaction involving immunologic mechanisms. It is initiated by production of specific antibodies, otherwise known as immunoglobulins, in reaction to food constituents. The body manufactures antibodies as part of its regular defense system against foreign invaders such as viruses and bacteria. In certain individuals, the immune system is triggered to elicit a specific antibody, called immunoglobulin E (IgE), against various environmental substances like pollens, pet dander, insect venoms, and foods. The common

immunologic mechanisms involved can cause food-allergic reactions to resemble allergic reactions to honeybee stings or penicillin.

Mechanism

In food-allergic individuals, IgE antibodies are produced against food components and circulate in the blood. Upon reaching certain cells, known as mast cells and basophils, the IgE becomes fixed to the cell surface and remains there. These cells contain high quantities of special receptors for IgE; rat mast cells have been found to contain 2.5×10^5 receptors per cell (Mendoza and Metzger 1976). A large portion of the IgE in the body is fixed to these cells, and when they become armed with IgE, they are said to be “sensitized.”

Mast cells are present in many tissues of the body, and basophils are present in the blood. When certain food components, or allergens, are ingested and circulate in the blood, they come in contact with the IgE bound to the mast cell or basophil surface. This contact causes a chain of events to occur in the cells, leading to the release of granules containing certain mediators that are responsible for the clinical manifestations of the allergic reaction. This immediate allergic reaction is classified as a Type I hypersensitivity reaction according to the Gell and Coombs classification (Coombs and Gell 1975). Most food-allergic reactions are of this type (Moneret-Vautrin 1986), although other mechanisms have been shown to be responsible in some cases (Chandra and Shah 1984).

The substances released from the cells include histamine and serotonin, but many other mediators are also released (Barrett and Metcalfe 1988). The major influences of these mediators are vasodilation, smooth muscle contraction, and a general increase in vascular permeability.

A small amount of any food protein remains undigested and passes through the gut wall and into the circulatory system (Steinmann, Wottge, and Muller-Ruchholz 1990). During the neonatal period the gastrointestinal barrier to protein uptake and transport is immature. Therefore, excessive amounts of food proteins may be transported into the circulatory system. In addition, it is possible that a large quantity of protein is not broken down intracellularly due to immature cell function. During this period, infants may become sensitized to ingested allergens; and later, when the mucosal barrier is mature, small amounts of allergen absorbed by the gut can cause a reaction.

It has been suggested that food-allergic reactions are caused by pathologically increased gut permeability (Sampson 1990a); this claim has been refuted by studies with cow's milk-allergic children in which the gut was not shown to be more permeable than in control children (Walker 1984; Powell et al. 1989). However, the permeability of the gastrointestinal barrier does increase following mast cell activation,

thereby allowing more allergen through the barrier and, perhaps, increasing the allergic response (Barrett and Metcalfe 1988).

Mast cells located on the mucous surfaces of the gastrointestinal tract are thought to play a major role in IgE-mediated food hypersensitivity reactions (Barrett and Metcalfe 1988). The gastrointestinal tract contains comparatively large numbers of mast cells – for example, the human duodenum contains 20,000 per cubic millimeter (Norris, Zamchek, and Gottlieb 1962), whereas human skin contains 7,000 per cubic millimeter (Mikhail and Miller-Milinska 1964).

Food allergy can develop only if the allergen crosses the gastrointestinal barrier. The clinical manifestations of food allergy depend on the transport of allergens across the barrier, giving rise either to local allergic reactions or, from the systemic circulation, to systemic responses. Penetration of the gut depends on the size and structure of the allergen, changes occurring as a result of digestion, gut permeability (depending on age and preexisting disease), and interaction with other antibodies located in the gut. More factors to be considered are individual responses to various amounts of allergen, sufficient IgE binding at the mast cell or basophil, and the susceptibility of the affected organs to the mediators that are released.

Symptoms

Clinical manifestations of mediator release depend on the part of the body that is affected. Table IV.E.3.1 lists symptoms associated with food-allergic reactions. The most common manifestations are gastrointestinal, dermal, and respiratory (Sampson and Cooke 1990), and they represent contact of allergen with populations of sensitized mast cells present in the affected tissues.

Gastrointestinal symptoms, including nausea and vomiting, cramping, abdominal distension, flatulence, and diarrhea, are especially frequent in infants and young children (Barrett and Metcalfe 1988). Chronic ingestion of offending foods by sensitive people may cause fecal blood loss, malabsorption, and other long-term manifestations (Buckley and Metcalfe 1982). Skin reactions to food allergens in such individuals can include pruritus, edema of the lips, tongue, gums, oral mucosa, and pharynx. Urticaria (hives) is the most common skin manifestation and can be diffuse or localized.

The reactions are very individualistic and diverse (Collins-Williams and Levy 1984). Initially, some individuals may experience immediate contact reactions on their lips and tongue, whereas others do not experience a reaction until the offending food has moved farther down the gastrointestinal tract.

Respiratory symptoms can be divided into two categories: upper airway distress, which in food-allergic reactions is usually caused by laryngeal edema, and middle airway distress, produced as a consequence of bronchoconstriction, with resulting edema and pulmonary mucus production (Collins-Williams and Levy 1984). Respiratory symptoms are infrequent with food allergies, although asthma has been associated with allergies to cow's milk, soybeans, and peanuts (Minford, MacDonald, and Littlewood 1982; Chiaramonte and Rao 1988).

Atopic dermatitis, a chronic inflammatory skin disease, is characterized by dry, easily irritated, intensely pruritic skin. The role of food hypersensitivity in atopic dermatitis has been debated for decades (Burks et al. 1988). IgE-mediated reactions to foods may contribute to the pathogenesis of atopic dermatitis (Sampson and McCaskill 1985); A. W. Burks and co-workers (1988) reported that 96 percent of atopic dermatitis subjects examined in their study developed skin reactions after food challenge.

Anaphylactic shock is a rare, acute, and potentially fatal form of food-allergic response. It is a generalized vasodilation triggered by significant mast cell degranulation and can involve a number of organ systems. If the cardiovascular system is affected, shock and low cardiac output result, and involvement of the respiratory tract can induce bronchospasm, pulmonary edema, and laryngeal edema. In food-induced anaphylactic shock, the gastrointestinal tract can respond with abdominal cramping and vomiting. Anaphylactic reactions may advance rapidly, beginning with mild symptoms that can progress to cardiorespiratory arrest and shock over a one- to three-hour period.

For example, an individual experiencing an anaphylactic food-allergic reaction may notice tongue itching and swelling, or palatal itching at first, then throat tightening, perhaps followed by wheezing and cyanosis. Chest pain and urticaria may be noted, and the individual may have gastrointestinal symptoms such as abdominal pain, vomiting, or diarrhea. A progression of symptoms can lead to potentially life-threatening hypotension and shock.

As there are no official means of reporting anaphylactic episodes, the actual frequency of these reactions is unknown, although J. W. Yunginger and colleagues at the Mayo Clinic (1988) documented 8 cases of fatal food-induced anaphylaxis in a period of 16 months. Exercise-induced food anaphylaxis is a type of reaction that develops during or shortly after exercise following the ingestion of certain foods. In documented cases, exercise most often occurred within 2 hours of ingestion. The foods implicated

Table IV.E.3.1. *Symptoms of food allergy*

Gastrointestinal	Respiratory	Cutaneous
Abdominal cramps	Asthma	Urticaria
Diarrhea	Laryngeal edema	Angioedema
Nausea	Rhinitis	Atopic dermatitis
Vomiting	Wheezing	

have included celery, lentils, peaches, shellfish, and wheat. Neither exercise alone nor food alone were sufficient to induce anaphylaxis (Maulitz, Pratt, and Schocket, 1979; Buchbinder et al. 1983; Kushimoto and Aoki 1985; Silverstein et al. 1986).

Symptoms of food-induced anaphylaxis may include cutaneous erythema, itching, and urticaria, which can progress to vascular collapse or upper respiratory obstruction (Schocket 1991). The pathophysiological mechanism of these reactions has not been elucidated, but heightened mast cell responsiveness to physical stimuli may be involved (Sheffer et al. 1983). It has been shown that avoidance of the offending food 8 to 12 hours prior to exercise usually eliminates difficulties.

Incidence and Natural History

The incidence of true food allergy is probably less than the general population perceives. Although studies have shown that at least one in four adults with allergies believe that they have experienced adverse reactions following ingestion or handling of foods (Sampson and Metcalfe 1992), it is estimated that just 2 to 3 percent (Bock 1987; Sampson and Cooke 1990) of the pediatric population and 1 to 2 percent of the adult population suffer from allergic reactions to food (Sampson and Cooke 1990). The true prevalence is unknown, however, and other estimates range from 0.3 to 10 percent of the total population (Taylor 1980; Johnstone 1981; Taylor 1985; Barrett and Metcalfe 1988; Schreiber and Walker 1989).

Food allergy is also influenced by culture and eating habits. For example, allergies to fish are more common in Japan and Norway than elsewhere because consumption of fish is higher in those countries (Aas 1966). The frequency of food hypersensitivity varies by ethnic group and socioeconomic class (Lieberman and Barnes 1990). Its etiology includes many factors, but genetics seem to play a large role. Studies with children have shown that the risk for allergy with one allergic parent is approximately 50 percent and for bilateral parental allergy 67 to 100 percent (Schatz et al. 1983).

Frequency of food allergy is highest in infancy and early childhood and diminishes with increasing age (Collins-Williams and Levy 1984). It is most prevalent between 1.5 and 3 years of age (Kjellman 1991). Although young children appear more likely to outgrow their food allergies, older children and adults may also lose their sensitivity if the offending food is removed from the diet. Investigations have shown that one-third of children and adults lose their clinical reactivity after one to two years of allergen avoidance (Sampson and Scanlon 1989).

Differences in disappearance rates depend on the allergen and the individual; for example, many children with allergy to cow's milk can tolerate small amounts of milk by the time they are 3 years old, and

egg allergy tends to decline before age 7. But allergies to nuts, legumes, fish, and shellfish tend to remain a problem for a much longer time (Bock 1982; Collins-Williams and Levy 1984).

Although exposure to food allergens can occur in utero through placental passage, most studies have shown that prenatal sensitization to food is infrequent (Halpern et al. 1973; Croner et al. 1982; Hamburger et al. 1983; Kjellman and Croner 1984). However, the case of an infant with strong skin-test reactivity to wheat and milk at the age of 5 hours has been reported (Kaufman 1971). Some fully breast-fed infants have been observed to have allergies to egg and cow's milk (Warner et al. 1980), and significant amounts of cow's milk casein have been detected in breast milk for up to 12 hours after maternal ingestion (Stuart et al. 1984).

There have been a number of studies on breast feeding to determine what benefit it might have in preventing food sensitization (Halpern et al. 1973; Saarinen et al. 1979; Juto and Bjorkstein 1980; Hide and Guyer 1981; Buscino et al. 1983). The results are conflicting, and the only consistent observation is that prolonged breast feeding (greater than four months and preferably six) and delayed introduction of solid food seem to be beneficial in the prevention of food allergy (Zeiger 1988). Breast feeding and maternal avoidance of common allergenic foods is suggested if there is a family history of allergy and a high serum IgE level at birth (Michel et al. 1980).

Causative Agents

Traditional Foods

The most commonly implicated foods in food allergy are listed in Table IV.E.3.2. The allergenic substances are usually proteins (Aas 1978b). Adults tend to be allergic to shellfish, legumes (especially peanuts), crustacea, tree nuts, and wheat (Bock and Atkins 1990; Sampson 1990a; Sampson and Cooke 1990). Children tend to be allergic to milk and eggs more frequently. Any food, however, has the potential to cause an IgE-mediated reaction. Very severe reactions are most often seen with peanuts, eggs, fish, and nuts (Collins-Williams and Levy 1984). In fact, peanuts are the most common cause of life-threatening reactions (Sampson 1990b). In a period of 16 months, a

Table IV.E.3.2. *Common allergenic foods*

Cow's milk
Crustacea – shrimp, crab, lobster
Eggs
Fish
Legumes – peanuts and soybeans
Shellfish – clams, oysters, scallops
Tree nuts
Wheat

research team at the Mayo Clinic reported 4 cases of death due to peanut-induced anaphylactic shock (Yunginger et al. 1988).

With the widespread introduction of new and "exotic" foods into our diets, it is anticipated that newly described allergic reactions will be reported. D. A. Moneret-Vautrin (1986), for example, has noted that with the relatively recent introduction of soybeans to the diet in France, the incidence of observed soybean allergy has increased to the point where it has become the third most common food allergy. Another example is kiwi fruit, which is not indigenous to the United States and became available as a result of importation in the 1980s. Following its introduction, allergic reactions to kiwi fruit were suddenly reported (Fine 1981; Fallier 1983).

Genetically Engineered Foods

Some concerns are being raised regarding nontraditional foods produced using genetic engineering techniques. These techniques allow for the fast and easy transfer of proteins from one food into another and for increased nutritional or other benefits, such as viral resistance. But there also is the potential for allergenic proteins from some foods to be transferred into other foods; how or whether these foods should be labeled is still being debated. Another compelling issue is whether genetic engineering techniques could potentially change the protein structure so that it could become allergenic once transferred and reproduced in the engineered food. The United States Food and Drug Administration has been and will continue to be working on policies governing the regulation of bioengineered foods, with allergenicity a key issue (Kessler et al. 1992).

Food Allergens

Most allergenic foods contain multiple allergens. For example, egg white is a complex mixture of at least 20 proteins, of which 5 or 6 are allergenic (Langeland 1982). The existence of multiple allergens in cow's milk (Goldman 1963; Bleumink and Young 1968) and peanuts (Barnett, Baldo, and Howden 1983) is well known. Most major food allergens range from 14 to 70 kilodaltons (1,000 atomic mass units) in molecular weight (Aas 1976; King 1976; Chiaramonte and Rao 1988). The upper molecular size observed for these allergens may be due to the permeability limits involved in the gastrointestinal barrier of the host. The allergen must be of adequate molecular intricacy to interact with the elements of the immune system, but quite low limits of molecular size exist; for example, small fragments of the codfish allergen, 6,500 and 8,500 daltons in size, possess pronounced allergenic activity (Elsayad et al. 1972). T. Ishizaka and K. Ishizaka (1984) have shown that the release of mast cell mediators depends on the bridging of two cell-fixed IgE molecules by the allergen. If such bridging is critical, the allergic response should, in fact, depend on the

number of allergenic determinants and their distribution on the surface of the allergen and not necessarily their size.

Common food allergens that have been fully or partially purified or characterized include codfish (Elsayad and Aas 1971), cow's milk (Goldman et al. 1963; Bleumink and Young 1968), eggs (Anet et al. 1985), peanuts (Barnett and Howden 1986; Burks et al. 1991; Burks et al. 1992), soybeans (Shibaski et al. 1980; Herian, Taylor, and Bush 1990), and shrimp (Hoffman, Day, and Miller 1981; Daul et al. 1992).

Most common allergenic food proteins appear to be heat-stable and resistant to proteolytic processes, such as peanut, shrimp, and codfish allergens, although many lose their allergenic activity during digestion or cooking (Aas 1984). Therefore, the allergenicity of a food may be influenced by the way in which the food is prepared, processed, or stored. For example, ovalbumin, the third most important allergen of eggs, is moderately heat-labile and is found only in small amounts in cooked egg (Hoffman 1979). Similarly, heating denatures some cow's milk allergens and makes them less allergenic (Bahna and Gandhi 1983). Individuals allergic to fresh fruits and vegetables can often tolerate these foods after they are cooked, canned, or frozen (Hannuksela and Lahti 1977; Eriksson 1978).

Cross-Reactivity

When a sensitive individual produces IgE antibodies directed against a certain food component and then encounters another food with similar components, an allergic response may occur. Structural allergenic similarities may exist within families of biologically related foods. Since dietary elimination of the offending food is the only "tried-and-true" method of preventing an allergic reaction, cross-allergenicity among related foods is of concern to allergists and patients. Food-allergic individuals often report cross-reactivity among other foods in the same families and are cautioned to avoid eating other closely related foods.

The most common food families associated with cross-reactions are fish, citrus fruits, legumes, shellfish, mollusks, and crustaceans (Chiaramonte and Rao 1988). Certain allergens are also apparently common to both foods and pollens. Common allergens have been reported in melon, banana, and ragweed pollen (Anderson et al. 1970), celery and mugwort pollen (Pauli et al. 1985), and apple and birch pollen (Lahti, Bjorkstein, and Hannuksela 1980).

"True" clinical cross-reactivity, however, is often difficult to establish without performing oral challenge studies. Reports in the literature seem to indicate that the occurrence of clinical cross-reaction is infrequent (Bock 1991). Extensive in vitro "allergenic cross-reactivity" (binding of IgE antibody that is specific for a different food allergen) has been documented. For example, D. Barnett, B. Bonham, and M. E. H. Howden

(1987) found that 25 percent of sera from legume-sensitive patients reacted strongly with peanut, garden pea, soybean, and chickpea extracts. Another research team also found extensive in vitro cross-allergenicity with peanuts, soybeans, peas, and lima beans in patients with legume sensitivity (Bernhisel-Broadbent, Taylor, and Sampson 1989).

However, an earlier study by the same researchers indicated that clinical results and in vitro results did not correlate well in evaluating allergenic cross-reactivity in the legume family (Bernhisel-Broadbent and Sampson 1989). Fifty-nine percent of skin-test-positive patients reacted to oral challenge, but only (approximately) 3 percent reacted in oral challenges to more than one legume. These studies show that although IgE antibodies can cross-react with related foods and cause positive skin tests in some cases, clinical manifestations are rare, and being allergic to one food does not necessarily rule out the consumption of biologically related foods.

Occupational Food Allergy

Development of allergic disease can be associated with occupational exposure to food proteins, and most occupational sensitizing agents are food-derived protein allergens (O'Neil and Lehrer 1991). Exposure is facilitated through inhalation and contact, and syndromes include occupational asthma, hypersensitivity pneumonitis, and skin reactions, including contact dermatitis and contact urticaria.

But although allergen levels in the workplace can be very high, only a small percentage of exposed workers develop occupational responses, suggesting that host factors play an influential role. For example, a history of allergy seems to influence development of sensitivity in some instances of occupational allergic disease, such as occupational asthma caused by exposure to green coffee beans (Jones et al. 1982). By contrast, however, family or individual history of allergy has not been shown to be a factor in the development of occupational asthma due to exposure to snow crab (Cartier et al. 1984).

Although the pathophysiology of many occupational food reactions has not been ascertained, the role of IgE in occupational asthma in bakers is well known (Hendrick, Davies, and Pepys 1976; Block et al. 1983). Most occupational allergens that are also foods have not been shown to induce symptoms following ingestion by workers sensitized via inhalation. Occupational asthma related to food products has been most often associated with seafood, eggs, grains, spices, herbs, coffee, tea, and mushrooms, whereas hypersensitivity pneumonitis, although low in occurrence, has been associated with molds used in cheese production, poultry proteins, fish meal, and mushroom spores. Dermatological reactions precipitated by food have been most often linked with fish, seafood, mustard, garlic, onion, cashews, and lettuce.

Diagnosis of Food Allergy

The diagnosis of food allergy is complicated. As we have seen, adverse food reactions can be caused by immunologic, nonimmunologic, and unknown factors (Hawrylko and Murali 1988). Substantiation of food allergy involves meticulous medical evaluation, since self-diagnosis is often unreliable. Elimination diets are frequently the initial approach to diagnosis of food allergy. The suspected offending food is first eliminated for two to four weeks, then reintroduced in small quantities and the patient's response is recorded (Cummings and Bock 1988).

Skin Testing

Skin testing is the most popular diagnostic tool for evaluating food allergy. It was first used in 1966 to identify children allergic to codfish (Aas 1966). The techniques for skin tests are varied and include puncture, prick, scratch, and intradermal methods. A positive response results in a skin reaction forming a "wheal-and-flare." Results are recorded by measuring the wheal-and-flare diameter and comparing them to negative (saline solution) and positive (histamine) controls. Clinically significant wheals are 3 millimeters or more in diameter (Cummings and Bock 1988). But the use of skin testing is precluded in the face of anaphylactic or other severe reactions, as well as generalized dermatitis, and positive skin tests may be inhibited by the use of some medications, especially certain antihistamines and drugs with antihistamine effects (Chodirker 1985).

A major problem with skin testing is that food allergen extracts have not been standardized. Food skin-test extracts are usually crude preparations of unknown composition, and accurate diagnosis of food allergy has, in part, been impeded by the lack of standardized extracts. This may ultimately be responsible for the variability in the predictive accuracy of food skin tests. H. A. Sampson (1988), for example, compared extracts from three commercial firms for their positive predictive accuracy against oral-food-challenge results and discovered that they ranged from 0 to 79 percent. Negative predictive accuracies were much better, ranging from 85 to 100 percent. C. Ortolani and co-workers (1989) found that skin-test extracts prepared with fresh foods produced more sensitive results when compared to commercial extracts. As an increasing number of food allergens are isolated and fully characterized, the potential for use of standardized food extracts increases.

Double-Blind, Placebo-Controlled Food Challenge

The double-blind, placebo-controlled food challenge (DBPCFC), in which a suspected offending food or placebo is administered in capsules or masked in other food, has been called the "gold standard" for the diagnosis of food allergy (Collins-Williams and Levy

1984; Cummings and Bock 1988; Sampson 1988; Bock 1991). This method was initially used in the investigation of immediate hypersensitivity reactions to foods in asthmatic children (May 1976). But S. A. Bock and F. M. Atkins (1990) recently reported a study of a group of food-allergic children over a period of 16 years. Fully 39 percent had positive DBPCFC results. Other investigators have reported good correlation of DBPCFC results with other laboratory methods for diagnosing food allergy (Bernstein, Day, and Welsh 1982; Sampson 1988; Parker et al. 1990).

Although controversy exists regarding the use of skin tests alone for diagnosis (Bernstein, Day, and Welsh 1982; Bock and Atkins 1990), some investigators have found good correlation between food-challenge studies and skin-test results. For example, one such study, which focused on a group of adults with histories of immediate adverse reactions to foods, discovered that 90 percent of those in the group who were food-challenge-positive were also skin-test-positive (Atkins, Steinberg, and Metcalfe 1985). Similarly, Bock and Atkins (1990) noted that 98 percent of positive food reactions to oral challenge were accompanied by a positive skin test in children past their third birthday. But in younger children the correlation dropped to 83 percent.

Nonetheless, other investigators have found false-positive skin tests in patients who were food-challenge-negative (Bernstein et al. 1982; Sampson and Albergo 1984). In general, a negative skin-test result is a good indicator of absence of IgE-mediated hypersensitivity (Sampson 1988; Parker et al. 1990).

Radioallergosorbent Assay

The first assay developed for measuring IgE blood levels was the radioallergosorbent assay (RAST), which was initially described in 1967 (Wide, Bennich, and Johansson 1967). RAST is specific and reproducible (Hawrylko and Murali 1988); however, it measures only circulating, and not tissue-bound, IgE. For this reason, it is less sensitive than skin tests for the detection of specific IgE. In addition, a RAST test is not necessary if the skin test is negative, as circulating IgE will not be present if there is no tissue-bound IgE (May 1976; Sampson and Albergo 1984; Chodirker 1985; Bock and Atkins 1990). Generally, sensitivity of RAST for food allergens varies from 50 percent to more than 90 percent (May 1976). RAST is still widely used in the allergy field to augment skin testing in diagnosis. Sensitivity of RAST for food allergens could be improved, as could skin-test reliability, if there were technical advances in the purification of food allergens.

Basophil Histamine Release

In this technique, blood basophils are obtained from an allergic patient. They are incubated with an allergen for a specified time, and the release of mediating histamine, caused by the presence of the allergen, is measured (May et al. 1970; Hirsch and Zastrow 1972).

In one study employing purified cow's milk proteins, wheat, soybean, and ovalbumin extracts, release of histamine from cells was found in 25 to 50 percent of pediatric subjects who also had positive skin tests, and no histamine was released in skin-test-negative patients (May and Alberto 1972).

In another study, however, food extracts did not induce release of histamine in patients with food allergy (Dockhorn 1969). H. Nolte and colleagues (1989), using gut mast cells obtained during biopsies of food-allergic patients, found that histamine release from intestinal mast cells could not be correlated with histamine release by blood basophils. Therefore, there are differences in the ability of food extracts to elicit a response in mast cells from different organs.

As with other diagnostic allergy assays, the sensitivity of the basophil histamine release assay could probably be improved by the use of purified food-allergen preparations. In addition, the technique is neither convenient nor easy to perform and, therefore, is not used widely for diagnosis.

Treatment and Prevention of Food Allergy

Elimination Diets

Elimination diets are the only proven effective therapy for food allergy, once the offending food is identified (Chodirker 1985; Schreiber and Walker 1989). In patients with severe reactions to foods, strict avoidance is indicated, which may include cross-reacting foods, prepared foods (unless one is absolutely certain of all ingredients), and the inhalation of aerosols of the implicated food (Rao and Bahna 1988).

Despite the apparent low incidence of clinical cross-reactivity, food-allergic persons are encouraged to use caution when eating a biologically related food for the first time. The recommendation is usually to avoid the offending food for a number of years, if not for the remainder of the patient's life. However, this advice may be adjusted for children, as they tend to outgrow certain food allergies, such as those to eggs and milk.

Avoidance

Food-allergic individuals must read food labels and ingredient listings diligently, as these usually reveal the presence of the offending ingredient. This task can be difficult, and a certain amount of "education" in reading labels is necessary, because particular terms used by food processors can camouflage the presence of allergenic proteins. "Natural flavorings" may include soybean and/or milk proteins, and "caramel flavoring" may contain milk proteins. "Hydrogenated vegetable protein" commonly contains soybean proteins but can be comprised of other types of vegetable protein. In addition, it can be difficult to identify and avoid some types of allergenic foods. An example is peanuts, which after pressing and deflavoring, are then reflavored and sold as other

types of nuts, such as almonds. Though they smell, taste, and look like tree nuts, these deflavored peanuts have been found to retain their allergenic qualities (Nordlee et al. 1981) and can pose a serious threat to unsuspecting peanut-allergic people.

Sensitive individuals may also be inadvertently exposed to foods contaminated with allergenic proteins during processing. In one documented case, for example, the inadequate cleaning of common equipment caused sunflower butter to become contaminated with peanut butter and elicited an allergic reaction in a peanut-sensitive patient (Yunginger et al. 1983).

Antihistamines

Antihistamines have been suggested for use in children with mild food allergies before mealtime (Bahna and Furukawa 1983). However, since antihistamines work by competitive inhibition (Furukawa 1988), they cannot completely prevent allergic reactions, and individuals with severe reactions should not rely on them.

Perspective on the Development of Allergy

Allergic disease is difficult to trace in archaeological or historical records. Evidence of foods or medicines used in the past is very limited (Lieberman and Barnes 1990), and the symptoms of allergy can mimic those of other diseases, thereby concealing accounts of allergic reactions in records and notes.

It has been theorized that allergy is a fairly recent development in the history of the human race. A prominent characteristic of parasitic infections in humans is an elevated level of IgE (Johansson, Benich, and Berg 1972). But because parasitic infections and allergic disease seem to be mutually exclusive (Merrett, Merrett, and Cookson 1976), allergy could be the result of an ecological transition from endemic parasitosis to improved sanitary conditions that eliminate or reduce parasitic exposure (Lieberman and Barnes 1990). An interesting discovery is a higher frequency of allergies among urban than rural residents in developing countries, but higher IgE levels in rural people than among urban asthmatics and individuals with allergies (Godfrey 1975; Merrett et al. 1976). The advent of modern, hygienic conditions and the elimination of parasitic infection could, therefore, be responsible for spawning production of IgE against other substances in our environment.

Susan L. Hefle

Bibliography

- Aas, K. 1966. Studies of hypersensitivity to fish. *International Archives of Allergy* 29: 346–63.
1976. Common characteristics of major allergens. In *Molecular and biological aspects of the acute allergic reaction*, ed. S. G. O. Johansson, K. Strandberg, and B. Ovnas, 3–19. New York.
- 1978a. Studies of hypersensitivity to fish – a clinical study. *International Archives of Allergy* 8: 39–50.
- 1978b. What makes an allergen an allergen. *Allergy* 33: 3–14.
1984. Antigens in food. *Nutrition Reviews* 42: 85–91.
- Anderson, L. B., Jr., E. M. Dreyfuss, J. Logan, et al. 1970. Melon and banana sensitivity coincident with ragweed pollenosis. *Journal of Allergy* 45: 310–19.
- Anet, J., J. F. Back, R. S. Baker, et al. 1985. Allergens in the white and yolk of hen's egg. A study of the IgE binding by egg proteins. *International Archives of Allergy and Applied Immunology* 77: 364–71.
- Atkins, F. M., S. S. Steinberg, and D. D. Metcalfe. 1985. Evaluation of immediate adverse reactions to foods in adult patients. I. Correlation of demographic, laboratory, and prick skin test data with response to controlled oral food challenge. *Journal of Allergy and Clinical Immunology* 75: 348–55.
- Bahna, S. L., and C. T. Furukawa. 1983. Food allergy: Diagnosis and treatment. *Annals of Allergy* 51: 574–80.
- Bahna, S. L., and M. D. Gandhi. 1983. Milk hypersensitivity. I. Pathogenesis and symptomatology. *Annals of Allergy* 50: 218–23.
- Barnett, D., B. A. Baldo, and M. E. H. Howden. 1983. Multiplicity of allergens in peanuts. *Journal of Allergy and Clinical Immunology* 72: 61–8.
- Barnett, D., B. Bonham, and M. E. H. Howden. 1987. Allergic cross-reactions among legume foods – an in vitro study. *Journal of Allergy and Clinical Immunology* 79: 433–8.
- Barnett, D., and M. E. H. Howden. 1986. Partial characterization of an allergenic glycoprotein from peanut (*Arachis hypogaea* L.). *Biochimica et Biophysica Acta* 882: 97–105.
- Barrett, K. E., and D. D. Metcalfe. 1988. Immunologic mechanisms in food allergy. In *Food allergy*, ed. L. T. Chiaramonte, A. T. Schneider, and F. Lifshitz, 23–43. New York.
- Bernhisel-Broadbent, J., and H. A. Sampson. 1989. Cross-allergenicity in the legume botanical family in children with food hypersensitivity. *Journal of Allergy and Clinical Immunology* 83: 435–40.
- Bernhisel-Broadbent, J., S. L. Taylor, and H. A. Sampson. 1989. Cross-allergenicity in the legume botanical family in children with food hypersensitivity. II. Laboratory correlates. *Journal of Allergy and Clinical Immunology* 84: 701–9.
- Bernstein, M., J. H. Day, and A. Welsh. 1982. Double-blind food challenge in the diagnosis of food sensitivity in the adult. *Journal of Allergy and Clinical Immunology* 70: 205–10.
- Bleumink, E., and E. Young. 1968. Identification of the atopic allergen in cow's milk. *International Archives of Allergy* 34: 521–43.
- Block, G., K. S. Tse, K. Kijek, et al. 1983. Baker's asthma; clinical and immunologic studies. *Clinical Allergy* 13: 359–70.
- Bock, S. A. 1982. Natural history of food sensitivity. *Journal of Allergy and Clinical Immunology* 69: 173–7.
1987. Prospective appraisal of complaints of adverse reaction to foods in children during the first three years of life. *Pediatrics* 79: 683–8.
1991. Oral challenge procedures. In *Food allergy: Adverse reactions to foods and food additives*, ed. D. D. Metcalfe, H. A. Sampson, and R. A. Simon, 81–95. Boston, Mass.
- Bock, S. A., and F. M. Atkins. 1990. Patterns of food hypersensitivity during sixteen years of double-blind, placebo-controlled food challenges. *Journal of Pediatrics* 117: 561–7.

- Buchbinder, E. M., K. J. Bloch, J. Moss, and T. E. Guiney. 1983. Food-dependent exercise-induced anaphylaxis. *Journal of the American Medical Association* 250: 2973-4.
- Buckley, R. H., and D. D. Metcalfe. 1982. Food allergy. *Journal of the American Medical Association* 20: 2627-31.
- Burks, A. W., S. A. Mallory, L. W. Williams, and M. A. Shirrell. 1988. Atopic dermatitis: Clinical relevance of food hypersensitivity reactions. *Journal of Pediatrics* 113: 447-51.
- Burks, A. W., L. W. Williams, C. Connaughton, et al. 1992. Identification and characterization of a second major peanut allergen, Ara h II, with use of the sera of patients with atopic dermatitis and positive peanut challenge. *Journal of Allergy and Clinical Immunology* 90: 962-9.
- Burks, A. W., L. W. Williams, R. M. Helm, et al. 1991. Identification of a major peanut allergen, Ara h I, in patients with atopic dermatitis and positive peanut challenges. *Journal of Allergy and Clinical Immunology* 88: 172-9.
- Buscino, L., F. Marchetti, G. Pellegrini, and R. Berlini. 1983. Predictive value of cord blood IgE levels in "at-risk" newborn babies and influence of type of feeding. *Clinical Allergy* 13: 503-8.
- Cartier, A., J.-L. Malo, F. Forest, et al. 1984. Occupational asthma in snow-crab processing workers. *Journal of Allergy and Clinical Immunology* 74: 261-9.
- Chandra, R. K., and A. Shah. 1984. Immunologic mechanisms. In *Food intolerance*, ed. R. K. Chandra, 55-66. New York.
- Chiaramonte, L. T., and Y. A. Rao. 1988. Common food allergens. In *Food allergy*, ed. L. T. Chiaramonte, A. T. Schneider, and F. Lifshitz, 89-106. New York.
- Chodirker, W. B. 1985. Diagnostic methods for assay of specific IgE antibodies in patients with suspected allergic disease. *Canadian Medical Association Journal* 132: 1370-1.
- Collins-Williams, C., and L. D. Levy. 1984. Allergy to food other than milk. In *Food intolerance*, ed. R. K. Chandra, 137-86. New York.
- Coombs, R. R. A., and P. G. H. Gell. 1975. Classification of allergic reactions responsible for clinical hypersensitivity and disease. In *Clinical aspects of immunology*, ed. P. G. H. Gell, R. R. A. Coombs, and P. J. Lachmann, 761-81. Oxford.
- Croner, S., N. M. Kjellman, B. Ericksson, and A. Roth. 1982. IgE screening in 1701 newborn infants and the development of atopic disease during infancy. *Archives of Disease in Childhood* 57: 364-8.
- Cummings, N. P., and S. A. Bock. 1988. Clinical methods for diagnosis. In *Food allergy*, ed. L. T. Chiaramonte, A. T. Schneider, and F. Lifshitz, 289-99. New York.
- Daul, C. B., M. Slattery, J. E. Morgan, and S. B. Lehrer. 1992. Identification of a common major crustacea allergen. *Journal of Allergy and Clinical Immunology* 89: 194.
- Dockhorn, R. J. 1969. Histamine release from the leukocytes of food-allergic individuals using whole food extract antigens. *Annals of Allergy* 27: 409-16.
- Elsayad, S., and K. Aas. 1971. Characterization of a major allergen (cod): Observations on effect of denaturation on the allergenic activity. *Journal of Allergy and Clinical Immunology* 47: 283-91.
- Elsayad, S., K. Aas, K. Sletten, and S. G. O. Johansson. 1972. Tryptic cleavage of a homogenous cod fish allergen and isolation of two active polypeptide fragments. *Immunochemistry* 9: 647-61.
- Eriksson, N. E. 1978. Food sensitivity reported by patients with asthma and hay fever. *Allergy* 33: 189-96.
- Fallier, C. J. 1983. Anaphylaxis to kiwi fruit and related "exotic" items. *Journal of Asthma* 20: 193-6.
- Fine, A. J. 1981. Hypersensitivity reaction to kiwi fruit (Chinese gooseberry, *Actinidia chinensis*). *Journal of Allergy and Clinical Immunology* 68: 235-7.
- Furukawa, C. T. 1988. Nondietary management of food allergy. In *Food allergy*, ed. L. T. Chiaramonte, A. T. Schneider, and F. Lifshitz, 365-75. New York.
- Godfrey, R. C. 1975. Asthma and IgE levels in rural and urban communities of the Gambia. *Clinical Allergy* 5: 201-7.
- Goldman, A. S., D. W. Anderson, W. A. Sellers, et al. 1963. Milk allergy. I. Oral challenge with milk and isolated milk proteins in allergic children. *Pediatrics* 32: 425-43.
- Halpern, S. R., W. A. Sellars, R. B. Johnson, et al. 1973. Development of childhood allergy in infants fed breast, soy, or cow milk. *Journal of Allergy and Clinical Immunology* 51: 139-51.
- Hamburger, R. N., S. Heller, M. Mellin, et al. 1983. Current status of the clinical and immunologic consequences of a prototype allergic disease program. *Annals of Allergy* 51: 281-90.
- Hannuksela, M., and A. Lahti. 1977. Immediate reaction to fruits and vegetables. *Contact Dermatitis* 3: 79-84.
- Hawrylko, E., and M. R. Murali. 1988. Immunologic tests for food allergy. In *Food allergy*, ed. L. T. Chiaramonte, A. T. Schneider, and F. Lifshitz, 301-20. New York.
- Hendrick, D. J., R. J. Davies, and J. Pepys. 1976. Baker's asthma. *Clinical Allergy* 6: 241-50.
- Herian, A., S. L. Taylor, and R. K. Bush. 1990. Identification of soybean allergens by immunoblotting with sera from soy-allergic adults. *International Archives of Allergy and Applied Immunology* 92: 192-8.
- Hide, D. W., and C. M. Guyer. 1981. Clinical manifestations of allergy related to breast and cow's milk feeding. *Archives of Disease in Childhood* 56: 172-5.
- Hirsch, S. R., and J. E. Zastrow. 1972. Basophil degranulation: A new method of observation and its correlation with skin testing. *Journal of Allergy and Clinical Immunology* 50: 338-47.
- Hoffman, D. 1979. Immunochemical identification of allergens in egg white. *Journal of Allergy and Clinical Immunology* 64: 159-63.
- Hoffman, D. R., E. D. Day, and J. S. Miller. 1981. The major heat stable allergen of shrimp. *Annals of Allergy* 47: 17-22.
- Ishizaka, T., and K. Ishizaka. 1984. Activation of mast cells for mediator release through IgE receptors. *Progress in Allergy* 34: 188-235.
- Johansson, S. G. O., H. Bennich, and T. Berg. 1972. The clinical significance of IgE. *Progress in Clinical Immunology* 1: 157.
- Johnstone, D. E. 1981. Current concepts in the natural history of allergic disease in children. *Respiration* 42: 34-41.
- Jones, R. N., J. M. Hughes, S. B. Lehrer, et al. 1982. Lung function consequences of exposure and hypersensitivity in workers who process green coffee beans. *American Review of Respiratory Disease* 125: 199-202.
- Juto, P., and B. Bjorkstein. 1980. Serum IgE in infants and influence of type of feeding. *Clinical Allergy* 10: 593-600.
- Kaufman, H. S. 1971. Allergy in the newborn: Skin test reactions confirmed by the Prausnitz-Kustner test at birth. *Clinical Allergy* 1: 363-7.
- Kessler, D. A., M. R. Taylor, J. H. Maryanski, et al. 1992. The safety of foods developed by biotechnology. *Science* 258: 1561-2.
- King, T. P. 1976. Chemical and biological properties of some atopic allergens. *Advances in Immunology* 23: 77-105.
- Kjellman, N. M. 1991. Natural history and prevention of food hypersensitivity. In *Food allergy: Adverse reactions to foods and food additives*, ed. D. D. Metcalfe, H. A. Sampson, and R. A. Simon, 319-31. Boston, Mass.

- Kjellman, N. M., and S. Croner. 1984. Cord blood IgE determination for allergy prediction – a follow-up to seven years of age in 1,651 children. *Annals of Allergy* 53: 167–71.
- Kushimoto, H., and T. Aoki. 1985. Masked type I wheat allergy. *Archives of Dermatology* 121: 355–60.
- Lahti, A., F. Björkstén, and M. Hannuksela. 1980. Allergy to birch pollen and apple, and cross-reactivity of the allergens studies with the RAST. *Allergy* 35: 297–300.
- Langeland, T. 1982. A clinical and immunological study of allergy to hen's egg white. III. Allergens in hen's egg white studied by crossed radioelectrophoresis (CRIE). *Allergy* 37: 521–30.
- Lieberman, L. S., and K. C. Barnes. 1990. Food allergies and adverse food reactions: An anthropological perspective. In *Food allergies and adverse reactions*, ed. J. E. Perkin, 207–232. Gaithersburg, Md.
- Lifshitz, F. 1988. Food intolerance. In *Food allergy*, ed. L. T. Chiamonte, A. T. Schneider, and F. Lifshitz, 3–21. New York.
- Maulitz, R. M., D. S. Pratt, and A. L. Schocket. 1979. Exercise-induced anaphylactic reaction to shellfish. *Journal of Allergy and Clinical Immunology* 63: 433–4.
- May, C. D. 1976. Objective clinical and laboratory studies of immediate hypersensitivity reactions to foods in asthmatic children. *Journal of Allergy and Clinical Immunology* 58: 500–15.
- May, C. D., and R. Alberto. 1972. In-vitro responses of leukocytes to food proteins in allergic and normal children: Lymphocyte stimulation and histamine release. *Clinical Allergy* 2: 335–44.
- May, C. D., M. Lyman, R. Alberto, and J. Cheng. 1970. Procedures for immunochemical study of histamine release from leukocytes with a small volume of blood. *Journal of Allergy* 46: 12–20.
- Mendoza, G., and H. Metzger. 1976. Distribution and valency of receptor for IgE on rodent mast cells and related tumor cells. *Nature* 264: 548–50.
- Merrett, T. G., J. Merrett, and J. B. Cookson. 1976. Allergy and parasites: The measurement of total and specific IgE levels in urban and rural communities in Rhodesia. *Clinical Allergy* 6: 131–4.
- Michel, F. B., J. Bousquet, P. Greillier, et al. 1980. Comparison of cord blood immunoglobulin E concentrations and maternal allergy for the prediction of atopic diseases in infancy. *Journal of Allergy and Clinical Immunology* 65: 422–30.
- Mikhail, G. R., and A. Miller-Milinska. 1964. Mast cell population in human skin. *Journal of Investigative Dermatology* 43: 249–54.
- Minford, A. M. B., A. MacDonald, and J. M. Littlewood. 1982. Food intolerance and food allergy in children: A review of 68 cases. *Archives of Disease in Childhood* 57: 742–7.
- Moneret-Vautrin, D. A. 1986. Food antigens and additives. *Journal of Allergy and Clinical Immunology* 78: 1039–45.
- Nolte, H., P. S. Skok, A. Kruse, and P. O. Schiotz. 1989. Histamine release from dispersed human intestinal mast cells. *Allergy* 44: 543–53.
- Nordlee, J. A., S. L. Taylor, R. T. Jones, and J. W. Yunginger. 1981. Allergenicity of various peanut products as determined by RAST inhibition. *Journal of Allergy and Clinical Immunology* 68: 376–82.
- Norris, H. T., N. Zamcheck, and L. Gottlieb. 1962. The presence and distribution of mast cells in the human GI tract at autopsy. *Gastroenterology* 44: 448–55.
- O'Neil, C. E., and S. B. Lehrer. 1991. Occupational reactions to food allergens. In *Food allergy: Adverse reactions to foods and food additives*, ed. D. D. Metcalfe, H. A. Sampson, and R. A. Simon, 207–36. Boston, Mass.
- Ortolani, C., M. Ispano, E. A. Pastroello, et al. 1989. Comparison of results of skin prick tests (with fresh foods and commercial food extracts) and RAST in 100 patients with oral allergy syndrome. *Journal of Allergy and Clinical Immunology* 83: 683–90.
- Parker, S. L., A. Leznoff, G. L. Sussman, et al. 1990. Characteristics of patients with food-related complaints. *Journal of Allergy and Clinical Immunology* 86: 503–11.
- Pauli, G., J. C. Bessot, A. Dietemann-Molard, et al. 1985. Celery sensitivity: Clinical and immunological correlations with pollen allergy. *Clinical Allergy* 15: 273–9.
- Powell, G. K., P. J. McDonald, G. J. Van Sickle, and R. M. Goldblum. 1989. Absorption of food protein antigen in infants with food protein-induced enterocolitis. *Digestive Diseases and Sciences* 34: 781–8.
- Rao, Y. A., and S. L. Bahna. 1988. Dietary management of food allergies. In *Food allergy*, ed. L. T. Chiamonte, A. T. Schneider, and F. Lifshitz, 351–64. New York.
- Saarinén, U. M., A. Backman, M. Kajossari, and M. Simes. 1979. Prolonged breast-feeding as prophylaxis for atopic disease. *Lancet* 2: 163–72.
- Sampson, H. A. 1988. Comparative study of commercial food antigen extracts for the diagnosis of food hypersensitivity. *Journal of Allergy and Clinical Immunology* 82: 718–26.
- 1990a. Food allergy. *Current Opinion in Immunology* 2: 542–7.
- 1990b. Peanut anaphylaxis. *Journal of Allergy and Clinical Immunology* 86: 1–3.
- Sampson, H. A., and R. Albergo. 1984. Comparison of results of skin-tests, RAST, and double-blind placebo-controlled food challenges in children with atopic dermatitis. *Journal of Allergy and Clinical Immunology* 74: 26–33.
- Sampson, H. A., and S. K. Cooke. 1990. Food allergy and the potential allergenicity-antigenicity of microparticulated egg and cow's milk proteins. *Journal of the American College of Nutrition* 9: 410–17.
- Sampson, H. A., and C. C. McCaskill. 1985. Food hypersensitivity and atopic dermatitis: Evaluation of 113 patients. *Journal of Pediatrics* 107: 669–75.
- Sampson, H. A., and D. D. Metcalfe. 1992. Food allergies. *Journal of the American Medical Association* 268: 2840–4.
- Sampson, H. A., and S. M. Scanlon. Natural history of food hypersensitivity in children with atopic dermatitis. *Journal of Pediatrics* 115: 23–7.
- Schatz, M., R. S. Zeiger, M. Mellon, and R. Porreco. 1983. Asthma and allergic diseases during pregnancy: Management of the mother and prevention in the child. In *Allergy: Principles and practice*, ed. E. Middleton, Jr., C. E. Reed, and E. F. Ellis, 935–86. St. Louis, Mo.
- Schocket, A. L. 1991. Exercise- and pressure-induced syndromes. In *Food allergy: Adverse reactions to foods and food additives*, ed. D. D. Metcalfe, H. A. Sampson, and R. A. Simon, 199–206. Boston, Mass.
- Schreiber, R. A., and W. A. Walker. 1989. Food allergy: Facts and fiction. *Mayo Clinic Proceedings* 64: 1381–91.
- Sheffer, A. L., N. A. Soter, E. R. McFadden, Jr., and K. F. Austen. 1983. Exercise-induced anaphylaxis: A distinct form of physical allergy. *Journal of Allergy and Clinical Immunology* 71: 311–16.
- Shibaski, M., S. Suzuki, H. Nemob, and T. Kurome. 1980. Allergenicity of major component protein of soybean. *International Archives of Allergy and Applied Immunology* 61: 441–8.
- Silverstein, S. R., D. A. Frommer, B. Dobozin, and P. Rosen.

1986. Celery-dependent exercise-induced anaphylaxis. *Journal of Emergency Medicine* 4: 195-9.
- Steinmann, J., H. U. Wottge, and W. Muller-Ruchholz. 1990. Immunogenicity testing of food proteins: In vitro and in vivo trials in rats. *International Archives of Allergy and Applied Immunology* 91: 62-5.
- Stuart, C. A., R. Twiselton, M. K. Nicholas, and D. W. Hide. 1984. Passage of cows' milk protein in breast milk. *Clinical Allergy* 14: 533-5.
- Taylor, S. L. 1980. Food allergy. *Journal of Food Protection* 43: 300-6.
1985. Food allergies. *Food Technology* 39: 98-105.
- Walker, W. A. 1984. Absorption of antigens and haptens. In *Food intolerance*, ed. R. K. Chandra, 17-54. New York.
- Warner, J. O. 1980. Food allergy in fully breast-fed infants. *Clinical Allergy* 10: 133-6.
- Wide, L., H. Bennich, and S. G. O. Johansson. 1967. Diagnosis of allergy by an in vitro test for allergen antibodies. *Lancet* 2: 1105-7.
- Yunginger, J. W., M. B. Gaurerke, R. T. Jones, et al. 1983. Use of radioimmunoassay to determine the nature, quantity, and source of allergenic contamination of sunflower butter. *Journal of Food Protection* 46: 625-9.
- Yunginger, J. W., K. G. Sweeney, W. Q. Sturner, et al. 1988. Fatal food-induced anaphylaxis. *Journal of the American Medical Association* 260: 1450-2.
- Zeiger, R. S. 1988. Prevention of food allergy. In *Food allergy*, ed. L. T. Chiamonte, A. T. Schneider, and F. Lifshitz, 329-50. New York.

IV.E.4 🐞 Food-Borne Infection

Numerous infectious diseases are acquired by the ingestion of contaminated food, milk, or water. Such illnesses have a worldwide distribution, although, predictably, the incidence is greatest in those countries where deficiencies exist in the provision of adequate sanitation and hygiene.

Yet outbreaks of food poisoning and other food-borne infections occur with frequency even in the most developed countries of the world. Here increasing demand for vast quantities of readily available food has been met by increased reliance on intensive farming techniques, industrialization of food production, introduction of various methods for preservation, and expanding networks for transport, storage, and distribution.

Many of the stages involved in the commercial processing and production of food provide ideal opportunities for the entry, maintenance, and multiplication of microbes. Consequently, several outbreaks of food-borne infection occur each year in developed nations, with the size of the outbreak reflecting the extent of the distribution of the product. Containment measures rely heavily upon early diagnosis and close cooperation between clinicians, medical microbiologists, and officials of public-health departments and are aimed at rapid detection of the source of infection and prompt action to curtail the outbreak. The mea-

sures taken must include immediate withdrawal of the incriminated article of food from sale to the public, dissemination of clearly worded information to consumers defining the source and mode of spread of the infection, and - where appropriate - guidelines on methods of preventing cross-infection between individuals.

Outbreaks of food-borne diseases occurring in Britain in recent years have included (1) salmonellosis spread via infected poultry meat, powdered baby milk, hens' eggs, imported chocolates, and spaghetti; (2) *Escherichia coli* infection transmitted through hamburger meat; (3) botulism from yoghurt that was flavored with contaminated hazelnut puree; and (4) listeriosis from soft cheeses. Recognition of the phenomenal cost entailed in the investigation and control of such outbreaks has led to a heightened awareness of the need for effective measures to curtail the incidence of food-borne diseases. Accordingly, recommendations have been formulated for improving the microbiological quality of food. However, the subject is a complex one and problems continue to arise.

Natural History of Infection

Food and water may transmit infectious diseases when they contain critical numbers of live pathogenic microorganisms that are adapted to invade and colonize the body of the victim through the gastrointestinal tract (see Table IV.E.4.1). Typhoid fever and viral hepatitis are typical examples of such diseases. In some cases the mere presence in food of products of microbial growth, such as toxins, is sufficient to induce disease even though the microbes themselves may have been destroyed. These microbial toxins are relatively heat-stable, retaining their activity even after the food has been heated before consumption, as, for example, staphylococcal enterotoxin and botulinum toxin. Strictly speaking, these are examples of food-borne intoxications in contrast to infections, but it is customary among microbiologists to include them under the rubric of infections.

A characteristic feature of food-borne infection is that, in general, the food concerned looks and tastes appetizing, with no obvious signs of spoilage. This is because the large numbers of bacteria required to cause symptoms in humans would also cause visible signs of spoilage in the food. And that food is unlikely to be consumed.

Thus, in most cases of food-borne infection, the causative agent does not come from spoiled food but from contaminated food and water that have been infected by traveling from host to host via the fecal-oral route. That is to say, the microbes discharged in the feces of an infected human or animal find their way via food, water, or unwashed hands into the mouth of a new host. The organisms are swallowed and proceed to establish themselves on, or in, the mucus membrane lining the gut, or they may pen-

Table IV.E.4.1. *Organisms causing food-borne disease*

Cestodes (Tapeworms)	<i>Taenia saginata</i> , <i>Taenia solium</i> <i>Diphyllobothrium latum</i> <i>Echinococcus granulosus</i> <i>Hymenolepis nana</i>
Nematodes (Roundworms)	<i>Trichinella spiralis</i> <i>Ascaris lumbricoides</i> <i>Enterobius vermicularis</i> <i>Trichuris trichiura</i> <i>Angiostrongylus</i> species <i>Gnathostoma spinigerum</i>
Trematodes (Flukes)	<i>Fasciola hepatica</i> <i>Fasciolopsis buski</i> <i>Opisthorchis</i> species <i>Heterophyes heterophyes</i> <i>Metagonimus yokogawai</i> <i>Gastrodiscoides hominis</i> <i>Paragonimus westermani</i>
Protozoa	<i>Cryptosporidium parvum</i> <i>Giardia lamblia/intestinalis</i> <i>Microsporidium</i> <i>Toxoplasma gondii</i> <i>Sarcocystis</i> species <i>Entamoeba histolytica</i> <i>Balantidium coli</i>
Fungus	<i>Aspergillus</i> and <i>Fusaria</i> species
Bacteria	<i>Campylobacter coli</i> , <i>Campylobacter jejuni</i> <i>Salmonella</i> enteric fever group: <i>Salmonella typhi</i> , <i>Salmonella paratyphi</i> A, <i>Salmonella</i> <i>paratyphi</i> B, <i>Salmonella paratyphi</i> C gastroenteritis group: <i>Salmonella enter-</i> <i>itidis</i> , <i>Salmonella typhimurium</i> , and others <i>Shigella dysenteriae</i> , <i>Shigella flexneri</i> , <i>Shigella boydi</i> , <i>Shigella sonnei</i> <i>Escherichia coli</i> : enteropathogenic (EPEC), enterotoxigenic (ETEC), enteroin- vasive (EIEC), enterohaemorrhagic (EHEC) <i>Yersinia enterocolitica</i> , <i>Yersinia pseudo-</i> <i>tuberculosis</i> <i>Vibrio cholerae</i> , <i>Vibrio eltor</i> , <i>Vibrio</i> <i>parahaemolyticus</i> , <i>Vibria vulnificus</i> <i>Clostridium welchii</i> , <i>Clostridium botulinum</i> <i>Bacillus cereus</i> , <i>Bacillus licheniformis</i> , <i>Bacillus subtilis</i> , <i>Baeillus anthracis</i> <i>Staphylococcus aureus</i> <i>Aeromonas hydrophila</i> <i>Plesiomonas shigelloides</i> <i>Brucella</i> : <i>Brucella abortus</i> , <i>Brucella</i> <i>melitensis</i> , <i>Brucella suis</i> <i>Mycobacterium bovis</i> <i>Coxiella burnetti</i>
Virus	Hepatitis A virus and hepatitis E virus Poliovirus Rotavirus Small round structured viruses (Norwalk and Norwalk-like agents) Astrovirus Calicivirus
Prion (?)	Agent causing bovine spongiform encephalopathy, which may infect humans who eat diseased meat (not proven but suspected)

trate the gut wall to invade other parts of the body, traveling in the bloodstream and lymphatic channels.

Multiplication of the organisms in their chosen sites is often followed by symptoms of the disease, although symptomless or subclinical infection may also occur. In either case, large numbers of the organism leave the patient (voided in feces, vomit, urine) ready to repeat the cycle in a new and susceptible host. Such diseases may belong to one of two categories: (1) diseases of the gastrointestinal tract, in which the symptoms may include diarrhea, vomiting, and abdominal pain, such as cholera, dysentery, or bacterial and viral gastroenteritis; and (2) diseases with symptoms not necessarily related to the gastrointestinal tract, such as toxoplasmosis, poliomyelitis, brucellosis, and typhoid fever.

Source of Infection

Infection at the Source

The meat of an animal infected with the parasite causing toxoplasmosis, sarcocystis, or tapeworm is infective when eaten undercooked or raw. Eggs laid by hens that are infected with *Salmonella enteritidis* often transmit infection when the egg is eaten raw or undercooked. The question of whether bovine spongiform encephalopathy may be readily transmitted to people eating beef from infected cattle is still hotly debated, but it is becoming more widely accepted that this transmission can occur. Milk from infected animals, when consumed without prior heat treatment, transmits diseases like tuberculosis, brucellosis, and Q fever.

Infected fish may also transmit infection (fish tapeworm) when eaten raw or undercooked. Shellfish contaminated with the organism *Vibrio parahaemolyticus*, which normally grows in the salty water of warm seas, cause food poisoning when ingested. Some fish and mollusks acquire powerful heat-stable toxins through their food and are then able to transmit diseases like ciguatera fish poisoning and paralytic or neurotoxic shellfish poisoning to people eating them. These animals feed on toxin-containing dinoflagellates and concentrate the toxins in their flesh. The fish, however, appear to be unharmed by the toxin. In contrast, scombroid fish poisoning is acquired by eating a preparation of the scombroid group of fish, like smoked mackerel, that has been subjected to substandard processing and/or improper storage conditions prior to smoking. Here, bacteria that are commonly present – such as species of *Proteus* and *Klebsiella* – act on, and break down, the histidine present in the musculature to yield histamine and other substances (scombrototoxin). The symptoms produced on ingestion mimic those of histamine poisoning.

Contamination in the Slaughterhouse

Meat is often contaminated with organisms that normally inhabit only the lumen of the gut of animals but

escape when the intestinal wall is damaged during slaughter and evisceration. This occurrence seems to be the most common route of transmission of such infections in Britain. Typical of this group of pathogens are *Clostridium perfringens* and species of *Salmonella*, *Campylobacter*, and *Yersinia*. Some species of *Salmonella* are able to cause a disseminated septicemic illness in their hosts, as, for example, *S. enteritidis* in poultry, which spreads to the heart and other organs that are used as giblets.

Food that is contaminated in this way may then contaminate other clean food by direct contact, as happens when raw meat comes into contact with cooked food, either directly as in refrigerators, or indirectly through use of the same preparation surface or knife.

Fecal Contamination of Food and Water

Both human and animal feces, which contaminate food and water, are major causes of diarrheal diseases. These diseases are especially prevalent in countries lacking adequate facilities for the safe disposal of sewage or protected water supplies. The role of flies, rodents, and birds in the transfer of pathogens from infected feces to food is well documented. Moreover, the washing of salad vegetables or fruit in water contaminated with human or animal sewage results in the transfer of diseases like typhoid, cholera, and bacillary or amoebic dysentery, as well as gastroenteritis caused by *E. coli*, *Campylobacter*, and *Salmonella* infections.

Bivalve shellfish (oysters, clams, and mussels) that are harvested from shallow inshore coastal waters adjacent to an outflow of sewage often contain large numbers of fecal pathogens. These animals are filter feeders, filtering large amounts of water each day to extract their food. The microbes present in fecal particles are concentrated in the flesh of the mollusk, which – when eaten raw or undercooked – may transmit diseases like typhoid fever, hepatitis A and E, and viral gastroenteritis. In addition, whereas the process of depuration (leaving the shellfish in tanks of purified water for over 24 hours prior to sale) removes most of the bacteria from the mollusks, viruses remain intact. Heat treatment is therefore required to inactivate them.

Waterborne infections commonly occur in areas lacking chlorinated water supplies. They are also associated with wars and other causes of sudden, uncontrolled movement of populations that disrupt normal services. Waterborne infections occur from time to time in developed countries, resulting in explosive outbreaks in the area of distribution of the water supply concerned.

Generally, such outbreaks follow a sudden or unforeseen breakdown of chlorination or filtration in the water purification plant, but they are also seen in institutions that are situated in remote areas and maintain their own water supply. It is notable that the

cysts of some gastrointestinal protozoan parasites, such as species of *Giardia*, along with *Entamoeba histolytica*, are resistant to the standard chlorination process used for drinking water. Improperly filtered but adequately chlorinated tap water has resulted in several recent outbreaks of cryptosporidiosis in Britain and the United States.

In developing countries, where infective diarrheal diseases are common, debility arising from the illness compounds the effects of malnutrition, and childhood mortality rates are excessive. Up to 15 out of every 1,000 children in some areas die before the age of 5 from diarrheal disease transmitted by water.

Food Handlers

Food may be contaminated due to poor standards of hygiene among those who handle and prepare it. These food handlers may be: (1) “carriers” who harbor and continue to excrete pathogenic microorganisms and are thereby able to transmit them to others, even if they themselves no longer suffer from symptoms of the disease or have never had any symptoms; (2) people who are not yet symptom-free following gastrointestinal illness; and (3) those who are caring for patients with such diseases. The danger lies in the transmission of microorganisms to food via unwashed hands and poor toilet hygiene, or by aerosols caused by sick food handlers vomiting in preparation areas. Food handlers typically cause outbreaks of staphylococcal food poisoning, viral gastroenteritis, and enteric fever, which is particularly significant in the commercial or bulk preparation of food. Staphylococcal food poisoning is especially relevant here since the toxin produced can resist heat treatment.

Contamination from the Environment

Microorganisms found in the environment may contaminate food and water. The spores (resistant forms) of *Clostridium botulinum* survive for long periods of time in the soil and mud of lakes and rivers. They may contaminate vegetables and fish and secrete a powerful neurotoxin into the food. *Listeria monocytogenes*, found in the gut of cattle, sheep, and other animals and in soil and grass, may contaminate foods such as unpasteurized milk products. Moreover, the ability of listeriae to grow and multiply at temperatures prevailing in most domestic refrigerators contributes to the problem of preventing infection.

Canned food has often been found contaminated with organisms that have gained entry through tiny undetected faults in the seal. During manufacture, cans of food are first heat-sterilized, then cooled by immersion in water before the labels are attached. When the cans undergo sudden cooling, material from the outside gets sucked into the interior, and infective agents gaining entry at this stage may then multiply freely in the food. If the food is to be eaten uncooked or only lightly cooked, it presents a hazard. A can of imported corned beef was implicated as the

source of the outbreak of enteric fever in Aberdeen in 1964. At the canning factory in South America, the cans were cooled in untreated water taken from a river close to the factory. Further investigation revealed that raw sewage from a neighboring settlement was discharged directly into the river upstream from the factory. Enteric fever was endemic in this community.

Back in Aberdeen, the infected corned beef was sliced and sold to the public through a retail outlet, resulting in the outbreak. The causative agent was identified and traced back to the retail premises, revealing the corned beef as the source of the outbreak. All cans belonging to the same batch as that causing the outbreak were withdrawn from sale. However, in spite of this action, the outbreak was observed to continue, with the diagnosis of more new cases in the area. It later became apparent that other items of precooked food sold in the shop, like ham, had been cross-contaminated when they were sliced on the same machine as the contaminated meat. Following this observation, the outbreak was brought under control by enforcing measures for (1) preventing further sale of contaminated food and (2) thorough cleaning and decontamination of all equipment and work surfaces involved. Canned food has also been implicated in outbreaks of botulism, staphylococcal food poisoning, and other diseases. In Great Britain, the most common food-borne infections today are campylobacter and salmonella gastroenteritis acquired from poultry and meat. Among the salmonellae, the incidence of *S. enteritidis* phage type 4 is now increasing steadily in England and many other countries.

Bacterial Food Poisoning

The food-poisoning group of organisms include *Staphylococcus aureus*, *Salmonella* species, *Clostridium perfringens*, *Bacillus cereus*, *Clostridium botulinum*, and *Vibrio parahaemolyticus*.

The term “food poisoning” is generally used in a restricted way to exclude clinical entities like typhoid and cholera, but it includes those conditions in which intestinal disturbances commonly occur, resulting in diarrhea and vomiting (with or without other symptoms). It is caused by ingesting either live bacteria or products of bacterial growth – the latter arising after the microbes have multiplied in the food for some hours before it is eaten.

This vital initial period of bacterial multiplication in food is an essential feature in the natural history of all bacterial food poisoning, as it ensures that a vastly increased number of bacteria are present in the food. The “challenge dose” of bacteria (the number that have to be ingested in order to cause disease) is much higher in the food-poisoning group than it is with other diseases, like typhoid and shigellosis, where food merely acts as a vehicle for transmission of the infection. *C. botulinum* is included in the food-poisoning

group, although the symptoms produced here are neurological rather than gastrointestinal, because this pathogen from the soil must grow and multiply for several hours in food before it is able to cause disease.

Crucial factors in the evolution of an outbreak of food poisoning include the type of food involved, the way in which it has been prepared, and the duration and temperature of storage prior to eating. As with most other food-borne infections, the food looks and tastes normal.

Bacterial food poisoning may belong to three different types, depending on the mechanism of pathogenesis observed:

1. In the *infective type*, as caused by *Salmonella*, large numbers of live organisms must be consumed. They then invade the cells lining the small intestine, where further multiplication occurs, causing inflammation, diarrhea, and vomiting. The time interval between ingesting the food and developing symptoms (the incubation period) is between 14 and 36 hours.
2. In the *toxic type*, caused by *S. aureus*, *B. cereus*, and *C. botulinum*, the phase of multiplication in food results in the release of toxins. When ingested, these toxins (preformed in the food) are responsible for producing the symptoms. It is not necessary to ingest the live organisms. Rapid heating of the food may destroy the bacteria but allow heat-stable toxins to remain active. In contrast to the previous type, the incubation period here is fairly short – a few hours – because the preformed toxin is able to act immediately on the gut.
3. The *intermediate type*, caused by *C. perfringens*, has an incubation period of 8 to 20 hours. The toxin is released not in the food but in the gut of the host, following ingestion. The release of toxins here coincides with the formation of spores (a resistant form of the bacterium). Prolonged low-temperature cooking or inadequate reheating of infected, leftover meat stews and casseroles increases the occurrence of this disease, because it stimulates sporulation of the *C. perfringens* in the gut of the host.

It is apparent that three important criteria have to be satisfied for the successful development of food poisoning. These are as follows:

1. Introduction of organisms into the food at some stage during its preparation.
2. Inadequate cooking. Most microbial agents in food are destroyed by cooking at high temperatures for the prescribed length of time. When frozen food, like meat, is cooked before it is completely thawed out, the temperature reached in the center of the meat may not be sufficient to kill the bacteria even when the outside looks well done.
3. Leaving food standing at an ambient temperature for several hours, enabling the bacteria to multiply

and attain the infective dose. Refrigeration of food stops multiplication of most of the harmful pathogens, with the exception of species like *Yersinia enterocolitica* and *L. monocytogenes*, which are particularly adapted for growth at temperatures of 0 to 4 degrees Celsius (°C). Freezing suspends replication of all pathogens but fails to kill them. After frozen food is allowed to thaw, bacteria may start to multiply again if kept standing at room temperature. Hence, careful attention is important at all stages of food preparation to ensure adequate cooking and prevent contamination. Furthermore, if food is not to be eaten without delay, it should be rapidly chilled before freezing for storage.

Clearly, even contaminated food is not harmful if properly cooked and eaten right away. But when eaten raw, inadequately cooked, or after remaining for a long time in a warm environment, contaminated food may transmit disease. Bacteria in food replicate rapidly by binary fission, doubling their numbers progressively at intervals of less than 20 minutes by a simple process of division in which each cell gives rise to two daughter cells. Food provides the ideal nutrient for bacteria, which continue to multiply, so that in just over 3 hours a starting population of 1,000 organisms (which is too low a dose to cause food poisoning) may reach the level of millions and constitute an effective challenge dose. In most outbreaks, food has been left for longer than 3 hours and the starting population may be higher than 1,000. Organisms like *Campylobacter*, *Shigella*, and the typhoid bacillus are infective at a much smaller dose level than the food-poisoning group, and in this case it is not necessary for the infected food to remain at room temperature for several hours in order to be able to transmit infection. Indeed, organisms like *Campylobacter* and viruses do not multiply in food. As already mentioned, in this case, food merely acts as a vehicle for transmission of these agents.

Most pathogenic bacteria are unable to grow in acidic food (pH less than 4.5) or in food with a low moisture content. Similarly, high salt or sugar concentration in preserved food inhibits many bacteria. However, they may survive for long periods of time in dried food products.

Staphylococcal Food Poisoning

The food-poisoning strains of *S. aureus* are usually harbored by human food handlers, either in septic skin lesions (boils and whitlows) or as a part of the normal resident microbial flora of the nose or skin. Transfer of such organisms to food may be restricted by the use of a no-touch technique in its preparation. This is especially relevant for large-scale processing of food. Food poisoning does not occur if the contaminated food is either eaten before the bacteria have a

chance to multiply, heated before the bacteria multiply and produce toxins (*S. aureus* are readily killed by heat), or refrigerated promptly and served chilled, to prevent multiplication of surviving bacteria.

If contaminated food is first refrigerated and then left standing for a considerable length of time in a warm room, the bacteria are able to multiply again. *S. aureus* multiplies rapidly over a wide range of temperatures, liberating its heat-stable enterotoxin. The type of food involved here is usually that which requires little or no cooking and has a high fat content, like cream cakes, trifles, and ham sandwiches. When eaten, the toxin acts on the host, causing symptoms of nausea, vomiting, and abdominal pain (but rarely with diarrhea) within 1 to 6 hours of ingestion. The toxin appears to act on nerve endings in the gut, transmitting its message to the center in the brain that controls vomiting.

Most cases are self-limiting. Recovery occurs rapidly, usually within 12 hours, and no treatment is required. If the infected food is heated to 70° C before serving, the *S. aureus* will be killed, but the toxin is still active, and even boiling for 30 to 60 minutes may not inactivate it. Moreover, *S. aureus* is not inhibited by a high concentration of salt. Therefore, foods like ham and other semi-preserved meat with a high salt and high fat content make an ideal medium for staphylococcal growth and multiplication. Prevention involves stringent measures for personal and environmental hygiene in the food preparation area, the covering of skin lesions with waterproof dressings, rapid refrigeration of prepared food that is not to be eaten right away, and thorough cooking of soups and meats before eating. Control of food handlers and use of the no-touch technique are important in the food manufacturing industry.

Salmonella Food Poisoning

Salmonellosis is a disease affecting both humans and animals and may be transmitted between them (a zoonosis). There are over 200 species of salmonellae, and they are widely distributed in nature, inhabiting the intestines of wild as well as domestic animals and household pets. They frequently contaminate meats, especially chicken, along with eggs, milk, and other products. Intensive farming techniques have resulted in widespread infection in the food chain, and mechanized mass processing of poultry carcasses ensures the cross-contamination of most poultry sold in shops.

Chickens infected with *S. enteritidis* are able to pass the infection on in their eggs. Infected humans acting as food handlers may also transmit infection unless scrupulous attention to hygiene is observed with regard to hand-washing after using the toilet and before handling food. Milk that is not heat-treated may also transmit salmonellae. The organism is readily killed by boiling for half an hour, so poor culinary practices are largely to blame for outbreaks of salmonellosis in humans. Examples of such practices

include cooking inadequately thawed large frozen joints of meat, low-temperature cooking, allowing cooked food to come in contact with uncooked meat or its juices, and prolonged storage at room temperature after inadequate cooking.

This last factor is the most important one because it enables an effective infective dose of salmonellae to develop in the food. Normal gastric acidity is sufficient to kill many ingested pathogens, but this defense may be breached when the microbe is taken in large numbers along with food that temporarily neutralizes the protective mechanism. Frequent outbreaks of salmonella food poisoning continue to occur in many developed nations because of (1) infected manufactured food; (2) the practice of adding raw egg to food that is to be eaten uncooked; and (3) importation of food products from areas with a high disease prevalence.

The strain of *Salmonella* responsible for an outbreak of gastroenteritis often points to the species of animal from which the food derived, as many strains have preferred hosts. *S. enteritidis* and *Salmonella typhimurium* are commonly found in poultry and cattle, *Salmonella dublin* in cattle, and *Salmonella badar* in turkeys. However, at times, certain strains appear to sweep through an entire population of livestock, probably because of contaminated feed and intensive farming techniques involving overcrowding and animal stress. The infection is often asymptomatic in humans and other animals, but continued excretion of the pathogen in feces enables transmission to occur.

This excretion is especially marked when animals are maintained in overcrowded situations and is reflected in the high incidence of infection in flocks of poultry and in beef cattle. During slaughter and evisceration, gut contents often contaminate the meat. Cross-contamination of other carcasses then occurs either through direct contact with the meat or by contact with contaminated surfaces and equipment. Thus, all retailed meat may be regarded as potentially contaminated.

Consumption of an infecting dose of 10^5 to 10^6 organisms is required in order to establish infection. This amount is readily achieved if improperly cooked meat is left to cool at room temperature for 3 or more hours. Fewer cells (10 to 100) may cause illness in very young children or in the elderly, especially if the pathogens are carried in high-fat foods such as cheese, cheesecake, salami, or hamburgers. Low numbers may also be infective when waterborne. A high fat content in food is believed to protect the salmonellae to some extent during cooking. This may be relevant to the survival of *S. enteritidis* phage type 4 in eggs and meat.

Since 1985, this particular strain has appeared with increasing frequency among the poultry population and has now spread to epidemic levels among them. In birds, the infection is not confined to the gut but

passes into the bloodstream to give a severe septicemia, and the eggs are infected before they are laid. Thorough cooking of the eggs or pasteurization of bulked eggs would destroy the pathogen. Pasteurization entails heating the product to a temperature of either 63 to 66° C for a period of 30 minutes or 71° C for 15 seconds. This form of heat treatment kills all vegetative pathogens like salmonellae in milk and eggs.

Human infection follows the ingestion of approximately 1 million organisms, which then invade the intestinal mucosal cells and cause local inflammation. The incubation period lasts for 14 to 36 hours and is followed by diarrhea and vomiting, with or without abdominal pain. Most infections are self-limiting and do not require any treatment, but in more severe cases, supportive therapy with rehydration is implemented. In very young babies or in elderly or debilitated patients, for instance, the disease may follow a septicemic course, giving rise to a serious life-threatening condition that necessitates the administration of appropriate antibiotics as well as rehydration. Following recovery, a small proportion of people may retain organisms in the gut for a varying period of time and continue to excrete them intermittently in the feces. If these “carriers” are employed as food handlers, they may transmit the organism to foods unless scrupulous attention to hygiene is observed.

Salmonella organisms may be introduced into manufactured food products at various stages in the production line. For example, they may enter after the heating stage and cause widespread outbreaks. Intercontinental outbreaks have been traced to chocolates and cheeses. Critical monitoring of production procedures to detect and control the possibility of microbiological contamination at every stage is of vital importance. Although it would be ideal to reduce salmonellosis in the livestock by measures like heat treatment of all animal feed and improvement of farming techniques, this aim is not readily achievable.

Legislation to enforce heat treatment of all milk sold from retail outlets has helped reduce milk-borne outbreaks in humans. However, the occasional undetected breakdown of a pasteurization unit has resulted in the sale of contaminated milk that has caused outbreaks. Countries lacking basic sanitation and hygiene tend to have a higher prevalence and transmission rate of salmonellosis. Foodstuffs imported from such areas (for example, spices like pepper) may require more than the usual decontamination. Alternatively, spices and the like can be added to food before it is cooked rather than afterward; this, too, prevents the problem.

***Clostridium Perfringens* Food Poisoning**

This food poisoning is associated with reheated leftover meat dishes, like stews. The causative organism is a normal commensal of the gut of humans and other

animals. It is a strict anaerobe able to grow and multiply only in an environment devoid of oxygen. Clostridia are able to produce tough, resistant spores that survive in adverse environments, and these spores germinate into vegetative cells. Multiplication can occur only in the vegetative state and not in spore form.

Inadequate cooking of meat enables the spores to survive. *C. perfringens* type A has particularly heat-resistant spores that can survive prolonged boiling, and these strains are associated with outbreaks of food poisoning, usually in institutions such as nursing homes. The type of food involved is, generally, meat dishes such as large joints cooked at a low temperature, or stews and broth when there is insufficient heat penetration to kill off all the spores.

Furthermore, the spores are protected by the protein-rich food. If the cooked food is eaten right away, it is generally safe, but when such food is left for between 3 and 5 hours at room temperature, it provides the ideal anaerobic conditions for the spores to germinate and the vegetative cells to multiply. Multiplication is rapid, with a mean doubling time of only 7 minutes at 41° C. The infecting dose of 10⁸ organisms per gram of food is soon attained. If, at this stage, the food is thoroughly reheated (to over 70° C), the vegetative cells are killed and the food is rendered safe.

If the food is only partially reheated (up to or below 60° C), though, the cells survive, and when ingested, they sporulate in the gut and shed the vegetative fragment, which releases a heat-labile enterotoxin. This acts on the small intestinal mucosa, damaging the brush border of epithelial cells at the tip of the villous processes. There is an outpouring of fluid and electrolytes resulting in violent, watery diarrhea and abdominal colic, occurring 8 to 20 hours after ingestion of the food. Vomiting and fever are uncommon, but nausea may be present. Recovery occurs within 24 hours in most cases, and no specific treatment is required. However, in elderly, debilitated individuals, severe dehydration may result, necessitating supportive therapy with rehydration.

Prevention involves the high-temperature cooking of meat in small amounts, rapid refrigeration if it is not eaten immediately, and reheating of all leftover meat dishes to a temperature higher than 70° C to destroy the vegetative cells. Cooking methods like frying and grilling have an advantage in that they deliver very high temperatures directly to contaminated meat surfaces.

***Bacillus Cereus* Food Poisoning**

B. cereus produces spores that help it to survive in adverse conditions such as drying and heat. Unlike the clostridia, this organism is an aerobe (growing in the presence of oxygen). It is found widely distributed as a saprophyte in the environment and on cereals like rice. In fact, outbreaks of *B. cereus* food poisoning are frequently associated with reheated rice

dishes such as fried rice and take-out fast food from Chinese restaurants. Bacterial spores survive the short cooking period involved, and if the rice is left at an ambient temperature for several hours to dry, germination of the spores occurs. The bacterial cells then multiply rapidly in the rice and liberate various enterotoxins. When the precooked rice is stir-fried to make fried rice, the heat-stable toxin is not destroyed by the short exposure to heat and may cause symptoms shortly after ingestion. When boiled rice is stored in a refrigerator, it goes lumpy, and when subsequently stir-fried, it fails to present the aesthetically acceptable appearance of well-separated grains of rice. This probably accounts for the tendency, especially in fast-food outlets, to leave precooked rice at room temperature for several hours and reheat it as required.

Two distinct types of disease symptoms are observed with *B. cereus* infection, depending on the type of toxin formed:

1. In the emetic type, vomiting is common. The incubation period here is 1 to 5 hours, and symptoms include nausea, vomiting, malaise, and, sometimes, diarrhea, lasting for 6 to 24 hours. This type is caused by a heat-stable toxin and is associated with cereals, especially the cooked rice just mentioned. The condition may, of course, be prevented by eating freshly cooked rice while it is hot; but if this is not possible, then the food should be maintained at over 60° C or stored at temperatures of about 8° C.
2. The diarrheal type is the rarer of the two and is due to the production of a heat-labile enterotoxin associated with lightly reheated meat sauces, puddings, and vegetables. Symptoms include profuse watery diarrhea, abdominal cramps, and, sometimes, vomiting occurring 8 to 16 hours after eating the food. Recovery is generally complete in 12 to 24 hours.

Botulism

Botulism is a neuromuscular disease caused by the ingestion of very small amounts of botulinum toxin. An oral dose of 0.005 µg may be lethal for humans. *C. botulinum* is a spore-producing organism that can grow only in the absence of oxygen. The tough spores are found worldwide in the soil and in the mud of lakes and rivers. When food like fish, vegetables, and meat is contaminated with soil containing the spores and is then left at room temperature, the spores germinate and multiplication of the organism occurs.

Some strains grow in temperatures as low as 3.3° C. During the growth process, a powerful neurotoxin is produced, which, when ingested, results in symptoms of botulism, following an incubation period of 18 to 36 hours. The preformed toxin causes a muscular paralysis with hoarseness, visual distur-

bances, headache, nausea, vomiting, and difficulty in swallowing and speaking. If left untreated, death from respiratory or cardiac arrest is a real danger. Symptoms may last from several weeks to many months.

A less severe disease is caused in infants ingesting spores of *C. botulinum*. These spores germinate in the infant's gut, releasing toxin in the intestine. Symptoms of flaccidity and weakness, with difficulty in feeding, are the presenting symptoms of infantile botulism. Treatment includes administration of antitoxin antibodies to attempt to neutralize the toxins.

Prevention relies on the destruction of spores by thorough heating of all preserved food. This is especially relevant to home-preserved food – the cause of many outbreaks. The high acid content of certain foods prevents the growth of this organism, but pre-formed toxin retains its activity when added to acid food. As mentioned previously, in an outbreak in England in 1989, the source was contaminated hazelnut puree used to flavor yoghurt. The toxin in food may be destroyed by heating at 85° C for 5 minutes.

***Vibrio Parahaemolyticus* Food Poisoning**

This form of food poisoning is related to the ingestion of undercooked or raw shellfish or fish. *V. parahaemolyticus* has a predilection for high salinity and grows well in the warmer seas. The shellfish and fish taken from such sites may be contaminated with the organism. When undercooked shellfish are eaten, symptoms of diarrhea and vomiting, fever, and abdominal cramps appear within 16 to 48 hours, resolving in 2 to 7 days. Other halophilic (salt-loving) bacteria, like *Vibrio vulnificus*, may cause a septicemic infection in people eating raw oysters. Prevention depends on the adequate cooking of seafood.

Other Bacterial Food-Borne Infections

Apart from the food-poisoning group of organisms, many other bacteria transmitted in food may also cause infection.

Campylobacteriosis

Campylobacteriosis occurs throughout the world and is the most common cause of acute infective diarrhea in most developed countries. The organisms *Campylobacter coli* and *Campylobacter jejuni* are widely distributed in nature, infecting the gut of domestic and wild animals, poultry and wild birds, and also humans. Poultry and cattle are important sources of human infection, because in addition to developing acute infection, they may become symptomless excretors of the organism with long-term carrier status. This is a classic example of a zoonosis (a disease transmitted between animals and humans) and, therefore, as is the case with nontyphoid salmonella, cannot easily be eradicated.

The curved, spiral-shaped organism is found in large numbers in the gut contents of infected animals

and is commonly transferred to poultry and animal meat during slaughter and evisceration. Most poultry carcasses sold in shops are consequently contaminated with the organisms. They may also be transferred from the feces of infected animals, wild birds, or humans to unprotected drinking water supplies or to food. Outbreaks occur most frequently after the consumption of improperly cooked poultry but have also taken place after the drinking of milk that was not heat-treated and unchlorinated water.

The infective dose is small (a few hundred bacteria), and the incubation period is approximately 3 days. The disease produced is an acute enterocolitis, and lesions are seen in the intestinal mucosa similar to those produced by shigellae or salmonellae. A characteristic feature of campylobacter is that the incubation period is followed by a period of a few days when the patient feels ill with headache, malaise, fever, and abdominal pain (the prodromal period). This is followed by a dysentery-like illness with frequent stools, sometimes containing blood and mucus. The prodromal symptoms help differentiate this illness from that caused by salmonellae or shigellae. Furthermore, the symptoms of abdominal pain and diarrhea seen here may be more prolonged. In fact, the pain may be so severe as to lead to a misdiagnosis of acute appendicitis and to the patient undergoing unnecessary abdominal surgery.

In most cases, recovery occurs within a week, and there is no need for antibiotic therapy. However, in debilitated or immunodeficient patients the organism may spread through the blood to cause a severe systemic infection, requiring appropriate antibiotic therapy.

Prevention of infection is achieved by measures that break the routes of transmission, such as the following:

1. Thorough cooking of all meat.
2. Observance of good principles of kitchen hygiene, like preventing contact between cooked food and raw meat, including contact with surfaces on which the raw food was placed first.
3. Heat treatment of milk by pasteurization or sterilization.
4. Adequate chlorination of the water supply or boiling of untreated drinking water.

Because the campylobacter organisms are so widespread in the animal kingdom, it would be unrealistic to hope to eradicate them entirely. It would, however, be useful to reduce the incidence of campylobacteriosis in poultry stocks and farm animals by good principles of farming and processing, while preventing human infection by observance of the simple measures recommended. This is also true of nontyphoid salmonellosis. Unlike salmonellae however, the campylobacter do not appear to multiply in food that is kept warm. The infective dose is low, and heavy contamination of food ensures transmission. Any food

may be contaminated by direct or indirect contact with the feces of infected domestic pets, other animals, birds, or humans.

Shigellosis (Bacillary Dysentery)

Shigellosis is a common disease seen throughout the world, but unlike nontyphoid salmonellae and campylobacter, shigellae infect only humans – there is no known animal host. The species involved are *Shigella dysenteriae*, *Shigella flexneri*, *Shigella boydi* and *Shigella sonnei*, and the severity of the diseases they produce decreases in this order.

Shigella sonnei, the mildest of the four forms, causes outbreaks mainly among young children in institutions like daycare centers and elementary schools. Poor toilet hygiene is the main cause of transmission in such cases. Outbreaks also occur in military barracks, institutions for the mentally handicapped, and in populations subjected to major social upheavals with a breakdown in the normal pattern of life.

Shigella dysenteriae, which causes the most severe of the four infections, is commonly seen in Asia. A high level of transmission is maintained in those countries with unprotected water supplies, a lack of facilities for the safe disposal of sewage and poor environmental conditions and personal hygiene practices.

The infective dose is small, and the disease is readily transmitted from person to person. There is no need for the organism to multiply in food in order to cause disease, as is the case with salmonella food poisoning. Following ingestion, the organism invades superficial layers of the cells lining the intestinal lumen, causing ulceration and inflammation that results in bleeding, secretion of mucus, and exudation of pus. The symptoms are diarrhea with loose stools containing blood, mucus, and pus. The severity of the illness varies with the species involved. The infection is localized to the intestine and does not invade deep tissues or the bloodstream except in very severe forms.

Antibiotic therapy is restricted to the serious forms of illness, like *S. dysenteriae*, or any severe form of dysentery occurring in the very young or debilitated patient. In most cases antibiotics are contraindicated as they may prolong symptoms and excretion of shigellae in the feces.

Prevention in the developed countries involves improving toilet hygiene, especially among children in institutions and schools. In developing countries lacking environmental hygiene and safe water supplies, travelers are well advised to drink only bottled or boiled water and to avoid uncooked or salad vegetables and fruit that is to be eaten unpeeled.

Typhoid Fever (Enteric Fever)

Typhoid fever is a severe febrile septicemic illness that is endemic in countries with poor sanitation and affects only human hosts – animals are not involved.

The causative organism, *Salmonella typhi*, is transmitted from one host to the next, usually through water supplies contaminated with human sewage. Food handlers also play an important role in the transmission. Recovery from this fever may be associated with the carrier state, which may last for a lifetime, with intermittent excretion of *S. typhi* in feces and urine. Such a state occurs when the organism continues to live in the gallbladder or kidneys of the host but no longer causes adverse symptoms. However, it is excreted from time to time in the feces of the carrier and may, therefore, be transmitted to other hosts. It is important to identify such carriers, educate them on food hygiene, and exclude them from following a career in the food industry.

Food may act as a vehicle for the transmission of this disease, which is by the fecal-oral route. Symptoms appear after an incubation period of about 10 to 14 days, with fever, headache, malaise, and, sometimes, diarrhea or constipation. Such symptoms are severe, and fatal complications may occur if untreated. Treatment with an appropriate antibiotic is important in all cases of enteric fever. Prevention requires the provision of protected, chlorinated water supplies and safe sewage disposal. In endemic areas it is wise to boil all drinking water or to treat it with chlorine tablets. Travelers to endemic zones are advised to drink only bottled water and not to eat salad vegetables, uncooked food, or unpeeled fruit. A vaccine containing heat-killed *S. typhi* is available but gives only partial protection, so the prevention of infection remains the main aim.

Paratyphoid fever caused by *Salmonella paratyphi* A, B, or C is a milder form of the infection just discussed. *Salmonella paratyphi* B is the strain transmitted in Europe, whereas *S. paratyphi* A and C are endemic in other areas.

Escherichia coli Gastroenteritis

This form of infection is possibly the most common cause of diarrhea in the world. Specific strains of this large family of organisms are involved in causing human gastroenteritis. Their natural habitat is the gut of humans and animals, and transmission is from person to person or by the intake of contaminated water or food. The infective dose is small, and there is no need for the organism to multiply in food for several hours before becoming infective.

The disease is prevalent in areas with poor environmental hygiene and frequently causes diarrhea in travelers to these areas (appropriately termed Aztec two-step, Rangoon runs, Delhi belly, Tokyo trots, and gippy tummy). In developed countries, outbreaks of infantile gastroenteritis have occurred in hospitals, and more recently, outbreaks of hemorrhagic colitis and hemolytic uremic syndrome have been reported where the causative organism was *E. coli* serotype 0157, transmitted in undercooked hamburgers. It is

possible that some strains cause zoonotic infection, whereas with others, human hosts are the sole source of infection.

The organisms have been classified according to the way in which disease is produced: Hence there is enteropathogenic *E. coli* (EPEC), enterotoxigenic *E. coli* (ETEC), enteroinvasive *E. coli* (EIEC), and verotoxin-producing or enterohemorrhagic *E. coli* (VTEC or EHEC).

In developing countries, diarrhea and dehydration brought on by such infection is a major cause of infant mortality. Adults are also affected from time to time, but they appear to develop a measure of immunity from constant exposure to all the different strains in the community. Travelers to infected areas, however, are unprotected and extremely susceptible to all the new organisms in the environment.

EPEC strains belong to certain known serogroups and, in the past, have caused outbreaks of gastroenteritis in infants and young children. Symptoms include diarrhea and vomiting, and severe dehydration may occur. Thus, unless supportive treatment is initiated rapidly, mortality rates can be high. Adults may also be occasionally infected. The mechanism of pathogenesis of EPEC strains is not clearly understood, but the organisms remain within the lumen of the gut and produce changes in the appearance of the cells lining the small intestine. Antibiotic therapy is restricted to the severely ill and debilitated cases.

All ETEC strains secrete enterotoxins, of which there are two types: heat-labile toxin (LT) and heat-stable toxin (ST). These toxins act locally on the cells of the intestinal mucosa, stimulating them in such a way as to result in a net outpouring of fluid and electrolytes from the bowel wall into the lumen.

The disease rarely occurs in temperate climates. In developing countries, however, it is common in children and is an important cause of travelers' diarrhea. Sewage contamination of water supplies is important in transmission. ETEC strains do not invade the body but remain in the gut lumen closely attached to the surface of the mucosal cells by means of short fibrillar processes (colonization factors). The toxin is secreted in the gut lumen and is internalized by the mucosal cells. Recovery occurs in a few days when the cells of the small intestine are replaced by fresh ones. Cells in the small intestinal mucosa are normally shed and replaced very rapidly. Because there is no inflammation in the gut wall, the stools contain no inflammatory cells. This is in sharp contrast to shigellosis and salmonellosis. Treatment is strictly supportive, with rapid correction of dehydration a lifesaving remedy.

EIEC strains (like *E. coli* serotype 0124) cause invasive disease similar to shigella dysentery, with ulceration of the intestinal mucosa that results in diarrhea with blood, mucus, and pus in stools. Food-borne infection is common, and cross-infection can occur.

EHEC (VTEC) strains have caused outbreaks of bloody diarrhea in people eating undercooked ham-

burgers. When meat is ground, it enables the organisms to grow better and cause greater contamination. The hemorrhagic colitis may be followed by hemolytic uremic syndrome (kidney failure and anemia due to the destruction of the red blood cells and destruction of thrombocytes). Many patients suffer only mild diarrhea, whereas others are seriously ill. *Escherichia coli* serotype 0157 is particularly associated with such outbreaks. The incubation period is 2 to 9 days. A contagious spread of the disease may also occur and is important in hospital outbreaks. Prevention is similar to the prevention of shigellosis with the added incentive to ensure that hamburgers are properly cooked.

Cholera

Cholera is a purely human, intestinal form of disease, characterized by profuse watery diarrhea (rice-water stools), that frequently leads to dehydration and death in untreated cases. It spreads along the fecal-oral route, with waterborne transmission most responsible for explosive outbreaks. Transmission from person to person also occurs. Flies can spread it, and the disease may also be food-borne. Fecal pollution of seawater can contaminate seafood which, if eaten raw or undercooked, may transmit the disease. No animal host has been detected for the cholera vibrio. Normal levels of gastric acid provide some defense against such infections in humans. The disease is more likely to occur in those with no or low levels of gastric acid.

Vibrio cholerae remains in the lumen of the gut attached to the surface of mucosal cells. The pathogenesis is mediated by a toxin that is secreted into the gut by the organism. This toxin is internalized by the cells and stimulates an outpouring of fluid and electrolytes, resulting in profuse diarrhea with the potential of causing death from dehydration (up to 20 liters per day of fluid loss). The mode of action of the cholera toxin is similar to that of *E. coli* LT. Indeed, the two toxins are similar in structure. There is no damage to the mucosa, and with modern medical treatment, recovery occurs in a week's time in most cases.

That treatment involves rapid rehydration. Oral rehydration with frequent small drinks of an electrolyte and glucose solution is usually used. In severe cases intravenous fluid infusion is necessary. Orally administered tetracycline is useful in reducing the period of excretion of the vibrio and also the severity of symptoms.

The latest cholera pandemic is due to the El Tor variant of *V. cholerae*, which causes a milder form of cholera than that seen in the nineteenth-century pandemics that swept the globe. It began in the 1960s in Southeast Asia and was spread by travelers, first to the Middle East and Africa and then to South America. Outbreaks, common where water supplies and environmental sanitation are poor or nonexistent, are especially likely in camps for refugees from wars or natural disasters.

Cholera prevention. Careful disposal of human sewage is essential, along with the provision of safe chlorinated water supplies or, failing that, boiled or bottled water. *Vibrio eltor* is more difficult to control because it causes a milder disease, meaning there are more ambulant cases to disseminate it more widely. A vaccine is available, but its protection is of short duration, and there is considerable doubt as to its efficacy.

Yersiniosis

Yersiniae are part of the normal gut flora of domestic and wild animals. Human infection is acquired by the ingestion of inadequately cooked meat containing sufficient numbers of the organism. *Y. enterocolitica* induces a typical food-poisoning-type syndrome with mild or moderate diarrhea, fever, and abdominal pain. Nausea and vomiting are rare, but mesenteric adenitis (inflammation of the abdominal lymph nodes) may occur and lead to misdiagnosis as acute appendicitis. Symptoms appear within 16 to 48 hours after the meal and may last for a period ranging from 24 hours to 4 weeks. Reactive arthritis is a complication that may occur following this infection, especially in people of HLA type B27. A heat-stable enterotoxin is produced by the organism that may contribute to the pathogenesis in the gut. However, the infection is not confined to the gut, as the organism invades the body to cause inflammation in the mesenteric lymph nodes.

A characteristic feature of this organism is its ability to grow and multiply in the cold at 0° C; consequently, refrigeration of contaminated food does not prevent the development of an infective dose. The bacteria, however, are readily inactivated by adequate cooking. Another strain, *Yersinia pseudotuberculosis*, is also found in the gut of animals and infects humans eating contaminated meat. It often causes a febrile illness with mesenteric adenitis, and sometimes a fulminant typhoid-like condition ensues. The yersiniae are associated with outbreaks of apparent acute appendicitis among young schoolchildren.

Listeriosis

The listeriae are pathogens of humans and animals that have the unusual property of being able to grow well at refrigerator temperatures as low as 4° C. They are widely distributed in the environment as well as in the guts of animals and humans. Contamination of food is therefore likely unless particular care is taken at all stages of food preparation and production to prevent contamination. The types of food associated with recent large outbreaks include unpasteurized milk products (like soft cheeses), cook-chilled chicken from supermarkets, and coleslaw made with cabbage that was fertilized by sheep manure. The organism frequently contaminates animal feeds, increasing its prevalence in the environment, but it does not grow at a low pH in the presence of organic acids.

When ingested by humans, listeriae may bring on a mild form of influenza-like illness, but in neonates, the elderly, and debilitated persons they may cause severe septicemia or meningitis. In pregnant women a mild influenza-like illness, followed by abortion or stillbirth, is often reported. Newborn babies may acquire the organism from the maternal birth passage and develop a fulminating septicemia or meningitis.

The organism is destroyed by the pasteurization of milk or the thorough cooking of food. Prevention of the contamination of food with environmental organisms is an important consideration.

Plesiomonas Shigelloides

This organism is another environmental bacterium that is widely distributed in nature and has been occasionally implicated in sporadic cases and outbreaks of diarrhea in many countries. It is found in freshwater (rivers and ponds), mud, fish, dogs, cats, and other animals. Adequate cooking of food destroys the organism.

Brucellosis (Undulant Fever)

Brucellosis is also called undulant fever because the disease often presents with bouts of fever alternating with afebrile periods. It is a zoonosis and is primarily a pathogen of goats, cattle, pigs, and camels. Humans acquire the infection by ingesting live brucellae in unpasteurized milk from infected animals, as well as through infective secretions from animals. Veterinary surgeons, farmers, and abattoir workers are especially at risk.

The main species of importance to humans are *Brucella abortus* from cattle, *Brucella melitensis* from goats, and *Brucella suis* from pigs. *Brucella melitensis* causes a more severe infection (Malta fever) than the other two species. The symptoms follow the ingestion of infected goats' milk or goats' cheese, and the condition is endemic in Europe, Africa, and the Far East. *Brucella suis* causes infection in America. In cattle, *B. abortus* causes abortion because of the organism's predilection for the placenta due to the high concentration of erythritol in bovine placenta and in the fetal fluids. Vigorous multiplication, therefore, occurs at this site, and the cow's milk will contain live brucellae, making it highly infectious, as are products made from it like butter, cheese, and cream.

In humans, the infection causes a protracted systemic illness. The organism is distributed in the bloodstream to various tissues, causing joint and muscle pains, headache, malaise, and depression. In many cases the disease runs a mild course, but in others a prolonged and disabling form of presentation is seen, especially when untreated.

Routine pasteurization of milk stops this mode of transmission so that only those in direct contact with infected animals remain at risk because the bacillus lives inside the cells of the host. Treatment with

antibiotics must be continued over a long period and repeated to ensure recovery.

Tuberculosis

Tuberculosis is generally transmitted by a respiratory route, but it can also be acquired by drinking unpasteurized cows' milk. The organism *Mycobacterium bovis* infects cows and passes in the milk to humans. It usually causes enlargement of lymph nodes in the neck or abdomen; fever, loss of weight, and can cause pulmonary symptoms. Treatment with several antibiotics is required. The infection may also spread in the bloodstream to involve other sites like meninges, bone, joints, and other organs.

Tuberculosis prevention. Pasteurization, boiling, or heat sterilization of milk renders it safe for use. Infected animals may be identified by tuberculin skin testing. They are then slaughtered to prevent spread to other herds. In humans, the BCG vaccine may be used to confer protection against this disease.

Q Fever

Q fever is an influenza-like febrile illness associated with patchy pneumonic consolidation of the lungs that humans acquire from animals by drinking untreated milk or by inhaling their infective secretions. The causative organism *Coxiella burnetii* has a worldwide distribution, affecting various domestic and wild animals and birds.

This organism is particularly resistant to desiccation and can survive heat-treatment temperatures only slightly less than those recommended for pasteurization of milk. The infection in humans may occasionally be severe, with complications like osteomyelitis, endocarditis, meningoencephalitis, and hepatitis. Treatment requires antibiotics, and prevention involves adequate pasteurization or heat treatment of all milk.

Food-Borne Viruses

Any virus transmitted by the fecal-oral route may be transmitted in food. Viruses can contaminate food at its source or during preparation. Mollusks and shellfish growing in coastal waters that are polluted with fecal material often contain human intestinal viruses. Oysters, clams, and mussels, eaten raw or incompletely cooked, can result in outbreaks of infection by the hepatitis A virus and small round structured viruses (SRSVs) such as those seen recently around the Gulf of Mexico. Outbreaks of virus infection acquired by drinking unchlorinated water have also been reported in other countries (poliomyelitis and hepatitis A and E). Chlorination of water prevents waterborne virus infection. Depuration of oysters and bivalve mollusks in clean water does not remove viruses as easily as it removes bacteria. Therefore, adequate cooking, by raising the temperature to 85 to

90° C and maintaining it there for between 1 and 5 minutes, is recommended.

Viral infections may also be transmitted to food by food handlers who are excreting the virus. A characteristic feature of these viral food-borne infections is the frequent occurrence of secondary cases following the outbreak. Here, person-to-person transmission occurs after the initial food-borne outbreak.

Viral Gastroenteritis

The viral agent most frequently found to cause food-borne outbreaks of gastroenteritis in Great Britain belongs to the group of SRSVs also known as Norwalk and Norwalk-like agents. SRSVs cause an acute short-lived attack in which vomiting is a prominent feature. The disease usually affects adults and is associated with outbreaks of gastroenteritis in restaurants, canteens, and institutions like hospitals. Food handlers are implicated in the transmission, especially when they have had a bout of projectile vomiting in the food preparation area. But any kind of food may also be involved in this form of transmission; raw oysters, for example, have been the source of outbreaks in recent years. Secondary cases usually follow the primary cases and are a result of transmission by airborne aerosols. The clinical picture presented is of a flulike illness with nausea, vomiting, diarrhea, and abdominal pain, following an incubation period of 1 to 2 days. Symptoms resolve within a further 1 to 2 days.

Control measures include scrupulous personal and environmental hygiene, the use of chlorine-based compounds to disinfect contaminated surfaces, and adequate cooking of shellfish.

Rotavirus

Rotavirus infection affects mainly schoolchildren, though adults may also be infected. It has a worldwide distribution, causing enteritis following a fecal-oral or airborne spread – usually in winter months. The incubation period is 48 to 72 hours. The symptoms presented are diarrhea and vomiting. The disease is acute and self-limiting, but moderate dehydration may result. Malnourished and debilitated children in developing countries may be severely affected. Treatment is mainly supportive (rehydration), and recovery is seen in a few days.

Viral Hepatitis

Hepatitis A virus. Hepatitis A affects mainly children and young adults in schools, colleges, and military camps. The virus is excreted in feces 7 to 10 days before the onset of symptoms and for a few days after that. Transmission occurs by the fecal-oral route during this period. Food handlers are important in this transmission since they are infectious before they become ill. Transmission also occurs by water, from flies, and in food, especially shellfish. The incubation period is from 3 to 6 weeks and is followed by symp-

toms of anorexia, fever, nausea, vomiting, and abdominal pain. Jaundice appears 3 to 10 days later, and the patient then starts to feel better. Prevention includes standard measures of hygiene and well-cooked food. A vaccine is now available to protect against hepatitis A infection. Normal human immunoglobulin, to provide passive protection over a period of 4 to 6 months, has been given to those traveling to endemic areas.

Hepatitis E virus. Waterborne outbreaks of hepatitis E virus infection (a recently recognized entity) have been reported from many countries. It is transmitted by the oral-fecal route, and food-borne infection (especially by shellfish) is a distinct possibility. Preventive measures include the provision of safe water supplies and protection of food from fecal contamination.

Poliomyelitis

Poliomyelitis is a disease affecting only the human host and is transmitted by the fecal-oral route, mainly through contaminated drinking water, although food may also act as a vehicle of infection. The infection usually causes only mild intestinal or systemic disturbances. Occasionally, however, a serious form of the illness is seen with involvement of the meninges, severe paralysis, and even death. In these cases recovery may leave permanent disabilities. A good vaccine is available to prevent the disease, which, it is hoped, will mean its eradication all over the world through a mass vaccination campaign. Meanwhile, improvement in environmental and personal hygiene reduces its transmission.

Bovine Spongiform Encephalopathy (BSE)

BSE is a recently recognized disease of cattle caused by a transmissible agent (prion) that resembles the agent of scrapie in sheep. Cattle develop spongiform areas in the brain, with vacuoles appearing in the gray matter. The infectious agent is extremely resistant to heating and standard sterilizing temperatures, and its incubation period is very long, ranging from 2 to 8 years. It is thought that the scrapie agent from infected sheep may have adapted itself to a new host (cattle), possibly as a result of the practice of feeding sheep protein to cattle. Whether BSE can be transmitted to humans eating infected meat has been a topic of debate for some time, but recent concern over outbreaks of "mad-cow disease" in Britain has prompted a wide belief that this is so. Although as yet unproven, BSE infection from food is suspected to be the cause of a number of human cases of a new type of the rare, deadly brain malady known as Creutzfeldt-Jakob disease.

BSE prevention. Precautions to be taken include the destruction of all BSE-infected animals and their milk, banning the feeding of ruminant-derived protein to cattle, and prohibiting the sale of bovine offal from

organs like the brain, intestine, and spinal cord, where the agent may replicate. In recent years, the British beef industry has been badly hurt by these and other efforts to control the disease.

Protozoal Infections

Amebiasis

Amebiasis is a disease with a worldwide distribution that occurs mainly where sanitation and hygiene are lacking. Transmission is mediated by the viable cysts of *Entamoeba histolytica*, which are tough-coated, rounded forms of the parasite, excreted in the feces of infected humans. The incubation period is between 2 and 6 weeks. Following ingestion of the cysts in contaminated water or food, excystation occurs in the intestines, and the liberated vegetative cells or trophozoites establish themselves in the large bowel. Ulcers are produced in the bowel wall, resulting in symptoms of amebic dysentery, namely diarrhea with blood and mucus in the stools. Infection may spread in the blood to cause abscesses in other organs like the liver, lung, and brain.

In many cases, noninvasive, symptomless infection continues for years, making carriers of people who excrete cysts in their feces and who are, thus, a potential threat to others, but these cases may convert to the symptomatic state when conditions are appropriate – as, for example, in the immunocompromised patient. The vegetative form, called trophozoite (feeding stage), is only occasionally seen in the feces (in diarrheal stool), is noninfective, and is rapidly destroyed in the environment. But the presence of a tough wall around the cyst (resting stage) renders it highly resistant to adverse environmental conditions. Cysts can survive for 2 to 4 weeks outside the host. They develop from trophozoites in the lumen of the gut and are passed out in the feces. Nuclear division and further development occurs within the cyst, even after it has been voided from the body.

Travelers often develop symptoms following ingestion of contaminated food or drink in countries where transmission is high. Oral or anal sex is also a possible route of transmission of this disease (for example, the "gay bowel syndrome" occurring among homosexual people). Prevention is achieved by the provision of good-quality water supplies or by boiling all drinking water (cysts are not killed by chlorine), safe disposal of sewage, screening food handlers to exclude cyst passers, and treatment of already infected individuals.

Cryptosporidiosis

Occurring in both humans and animals, infection by *Cryptosporidium parvum* is transmitted to the former mostly in rural areas, following contact with animals or consumption of contaminated milk, water, or food that has not been heat-treated. Explosive outbreaks in urban areas (Milwaukee, for example,

where in 1993 more than 400,000 people became infected) are usually the result of waterborne spread. The cysts are not destroyed by chlorination but may be removed from drinking water by filtration. Indeed, a malfunction in the filtration of water in sand beds is believed to have caused several outbreaks of this disease in Great Britain. Outbreaks have also arisen from consumption of unpasteurized milk. The pasteurization process destroys this pathogen.

The incubation period is between 3 and 11 days. Following ingestion, the organism emerges from the cyst and penetrates the epithelial cells lining the intestinal lumen, coming to lie just beneath the membrane within a vacuole. Here, organisms multiply to form the oocysts that are released in the feces. The symptoms include impaired digestion, malabsorption, profuse watery diarrhea, vomiting, and weight loss. In the immunocompetent host, recovery occurs within 7 days. Immunosuppressed patients, however, develop severe and protracted diarrhea with dehydration and prostration, as is seen when AIDS patients are infected.

There is no effective treatment for this infection apart from supportive measures like rehydration. Prevention of infection involves preventing oocysts from coming in contact with food or water, provision of safe drinking water, and proper cooking of all food.

Giardiasis

Giardiasis is an infection principally of the small bowel, with symptoms of diarrhea, passage of bulky, pale, offensive stools, abdominal cramps and distension, malabsorption, weight loss, and fatigue. The malabsorption involves mainly fats and fat-soluble vitamins. The disease is characteristically seen as severe bloodless diarrhea in children. The trophozoite form of *Giardia lamblia/intestinalis* inhabits the duodenum and upper ileum, closely attached to the mucosa. Transmission is by the cyst forms, which are passed in stools that contaminate water and, sometimes, food. Cysts are destroyed by boiling but not by the routine chlorination of drinking water. Hence, the importance, as with cryptosporidium, of providing efficient preliminary sand filtration. The infective dose is said to be as low as 10 cysts. Following ingestion, excystation occurs, and the trophozoites are responsible for the intestinal pathology. Heavy infestation damages the mucosa and atrophies the villous processes, but there is no penetration into intestinal cells. The disease has a worldwide distribution and is found in relation to poor hygiene and the substandard quality of drinking water. The incubation period is between 1 and 3 weeks, and the disease, if untreated, lasts for 4 to 6 weeks. Symptomless carriage is a common feature. Prevention involves the use of boiled or filtered water and protection of food from contamination.

Other Intestinal Protozoa

Balantidium coli. This is a normal inhabitant (commensal organism) of the pig bowel. It has a worldwide distribution but rarely causes disease in humans. Those ingesting the cysts may develop a symptomless carrier status, or invasion of the colonic mucosa may lead to symptoms of colitis.

Isospora belli. This infection is particularly found in association with diarrhea in patients with acquired immunodeficiency disease (AIDS). In normal individuals it is possibly a cause of symptom-free infection or of mild transient diarrhea. Little is known about the life cycle. The organism inhabits the small bowel, cysts are passed in the stools, and transmission is believed to be by direct person-to-person transfer or via food or water.

Microsporidium. This group of organisms is found in a range of host species. The organisms have also occasionally been isolated from patients with AIDS and are now believed to be opportunistic pathogens. They are possibly transmitted during the spore stage, voided in the stools of an infected host, and they contaminate food or water. When swallowed by the new host, the organisms enter the gut mucosal cells and multiply there, producing symptoms of intractable diarrhea.

Toxoplasma gondii

Toxoplasmosis is a food-borne systemic infection with a wide geographical distribution. Serological tests indicate that a large proportion of the human population has been infected at some stage in life.

Toxoplasma gondii has a complex life cycle, with its definitive or final host (in which a sexual stage of development occurs) being the domestic cat and some of its wild relatives. The sexual stage takes place in the intestine of the cat, which then excretes oocysts in its feces, although the cat is rarely ill with the infection. Indeed, this relationship typifies a successful parasitization, where neither the host nor the parasite appears to suffer. The intermediate host, in which the asexual life cycle occurs, may belong to a wide range of species, including *Homo sapiens*, who is infected by ingesting the oocysts from cat feces. The parasites then travel via the bloodstream to infect cells in diverse parts of the body, where they multiply asexually (by fission), producing infective cysts. The incubation period may range from 5 to 20 days.

In the healthy individual, the infection is controlled by the immune response, although the parasite may evade immune surveillance by encysting in various tissue cells, surrounding itself with a tough coat, and continuing to multiply at a slow rate within the cyst. Damage to tissues may occur at these sites in immunocompromised individuals like AIDS patients. Pregnant women acquiring the acute infection may transfer the infection transplacentally to the fetus,

which may be seriously damaged as a consequence. As in the immunocompromised patient, lesions may occur in the fetal brain, lung, liver, and other tissues with a fatal outcome.

The majority of people infected (often in childhood) develop no symptoms, but some develop a glandular fever-like syndrome, with fever and generalized enlargement of lymph glands. The disease may sometimes present with inflammatory lesions of the eye (choroidoretinitis). Avoidance of infection involves rigorous attention to hygiene, prevention of contamination of food with cat feces, the cooking of all meat before consumption, and advice to pregnant women to avoid contact with cat litter and feces.

Sarcocysts

Sarcocystis hominis causes a disease in humans (the definitive host), with mild symptoms like nausea, diarrhea, and malaise appearing a few hours after the ingestion of raw, infected meat from animals like cattle, sheep, and pigs (intermediate hosts) that contain the encysted infective forms. It is a common infection in animals but rare in humans. This is perhaps because of low transmissibility, inactivation by cooking or freezing, or the frequent occurrence of subclinical infection. The oocysts are passed in the feces of infected humans, and animals ingesting these oocysts develop the tissue cysts.

Trematodes (Flukes)

Adult trematode worms may inhabit either the intestinal tract, the bile ducts, or the lungs. Trematodes are flat, leaflike organisms having an alternating sexual cycle in the final (definitive) host and an asexual multiplication cycle in the intermediate host. Many trematodes have, in addition, a second intermediate host.

Fascioliasis

The liver fluke, *Fasciola hepatica*, is a common parasite of sheep and cattle kept in damp pastures in many parts of the world, and it is only occasionally found in humans. The adults live in the bile duct of the host and produce eggs that are passed in the feces of sheep and cattle. The developing form (miracidium) leaves the eggs and enters a suitable snail host. Here it multiplies asexually, and in 4 to 5 weeks the infective form leaves the snail and encysts on vegetation such as grass, watercress, or radishes growing at the site. Humans acquire the infection by eating the raw, infected vegetables. The young fluke then excysts and penetrates through the duodenum, goes to the liver, and settles in the bile ducts.

Symptoms of dyspepsia, nausea, fever, and abdominal pain occur in the acute stage. In the chronic stage, pain, enlargement of the liver and obstructive jaundice, and anemia may occur. Prevention involves protecting vegetation to be eaten raw from cattle or

sheep feces and eliminating snails by draining pastures or using molluscicides.

Fasciolopsis

Fasciolopsis is the intestinal infection involving snails that occurs in China and some other parts of Asia. The large-sized fluke *Fasciolopsis buski* attaches itself to the wall of the small intestine of its host and passes its eggs in the feces. When the eggs find themselves in freshwater ponds, they hatch, and the larvae infect freshwater snails, in which they develop further. On leaving the snails, they encyst on the fruit and roots of water plants, like the water caltrop, water chestnut, and lotus. Pigs and humans are common hosts of the infection, which they acquire by eating the infective encysted forms on raw aquatic vegetation. Children in endemic areas often peel this material with their teeth, thus enabling entry of the parasite. Following ingestion, the parasite emerges from the cyst, attaches itself to the wall of the intestine, and develops into an adult.

In most cases the infection is asymptomatic, but where there is heavy infestation, symptoms include diarrhea, vomiting, and gripping pain. Diagnosis is confirmed by finding the eggs in the feces. Prevention involves careful peeling or cooking of water-grown vegetables before consumption in endemic areas. The cysts are killed by the drying of water plants or by a few minutes of boiling. It is necessary to keep human or pig feces away from ponds.

Heterophyiasis and Metagonimiasis

Heterophyes and *Metagonimus* species are small flukes that attach themselves to the wall of the small intestine between the villi. Eggs are passed in the feces into water, and larvae emerge to enter freshwater snails. The larvae emerge from the snails in due course to infect freshwater fish. Human infection follows ingestion of infective cysts in raw fish. The larvae excyst and establish themselves in the gut where they grow to adults, causing nausea, diarrhea, and abdominal colic. Adequate cooking of fish and prevention of fecal pollution of freshwater are important preventive measures.

Clonorchiasis and Opisthorchiasis

Infection with *Clonorchis* species occurs mainly in the Far East. Infection with *Opisthorchis* species occurs in the Far East, in Poland, in the Ukraine, and in Siberia. Adults are small flukes that live in the bile duct. Eggs, passed in the feces, infect snails in slow-flowing streams.

After a stage of development, larvae leave the snail and encyst themselves beneath the scales of freshwater fish. When raw or undercooked fish is eaten, the flukes excyst in the intestine and migrate up to the bile duct. Light infection may cause only vague symptoms related to the biliary tract. However, signs of biliary obstruction, portal hyper-

tension, and enlargement of liver and spleen may occur in heavy infection. Dogs can also be infected and act as reservoirs of infection. Prevention involves the cooking of fish and treatment of feces before their use as fertilizers.

Gastrodisciasis

Human infection occurs in India (Assam), Bangladesh, Vietnam, China, and Russia, where adult *Gastrodiscoides hominis* attach themselves to the large intestine (colon or caecum), and eggs pass out in the feces to enter and develop in pond snails. The infective form then leaves to encyst on vegetation, like water calthrop. Eating infected raw vegetables leads to human infection and superficial inflammation of the intestinal mucosa. The presenting symptoms are diarrhea and passage of mucus in stools. Pigs act as a reservoir of infection. Prevention includes the careful peeling and washing of all water vegetables and keeping pig feces away from the water. Dried water-plants can be fed to pigs because drying inactivates the parasite.

Paragonimiasis

Paragonimiasis is seen in Asia, Africa, and America. The lung fluke - *Paragonimus westermani* - infects humans and animals that eat crustaceans like freshwater crabs and crayfish. The fleshy adults live in pairs in the lungs, resembling large coffee beans in size and shape. Eggs may be coughed up in sputum or swallowed and passed in feces. When they reach fresh water, the larvae emerge and penetrate snails. After a stage of development, the larvae leave the snails to penetrate and encyst in crabs and other crustaceans. When these are eaten uncooked, they can transmit the infection. Pickling processes that fail to kill the larvae are also often involved in transmission of the infection. Excystation in the small intestine enables the young fluke to emerge, penetrate the gut wall, and migrate across the diaphragm and into the lungs. Granulomatous changes occur in the lung surrounding the fluke, and symptoms include dry cough, pain in the chest, difficulty in breathing, and the coughing up of blood. Prevention involves the proper cooking of crustaceans and care in handling and washing them so as not to cause cross-contamination with other food.

Cestodes (Tapeworms)

Adult tapeworms all have flat, tape-like bodies and live in the intestinal tract. The body is segmented and the eggs are contained in the terminal segments, which are passed out in the feces. These eggs are infective.

Taeniasis

Taenia saginata (beef tapeworm). Human infection with this tapeworm occurs when raw or slightly cooked beef is eaten. Cattle are infected by eating

infective eggs and develop cysticercosis, whereby infective larvae are found encysted in the muscles. When beef is eaten raw by humans, the larvae emerge and attach themselves to the walls of the small intestines and grow into adults of 5 to 20 meters in length. The terminal segments containing eggs drop off, emerging with feces to remain infective on grass and vegetation contaminated by human feces. The eggs are eaten by cattle and hatch out in the duodenum. The parasites then penetrate through the intestinal wall to reach the bloodstream and are carried to the muscles, where they form fluid-filled cysts (cysticerci). Each of these contains an infective larva that, when ingested raw by humans, can grow into an adult tapeworm in the gut. Cysts are killed by cooking at over 56° C.

The symptoms involved are vague abdominal complaints, with segments of worm occasionally wriggling out of the anus. Prevention includes cooking beef adequately, inspecting all meat intended for sale, and the condemnation of any “measly” meat containing tiny pinhead-sized cysts.

Taenia solium (pork tapeworm). People are at risk of contracting this tapeworm in countries where much pork is eaten, including Eastern Europe, China, Indonesia, South Africa, Mexico, and Chile. The infection is carried in pigs as infective cysticerci in the muscles. People acquire the infection by ingesting undercooked or raw pork. The infective form then leaves the cyst and attaches itself to the mucosa of the small intestine, where it proceeds to grow over the next 10 weeks to its full length of between 2 and 10 meters. The terminal segments containing the eggs are passed out in the feces. Pigs eat the infective eggs in soil or feces, and the larvae hatch out, penetrate the gut wall, and travel by the bloodstream to encyst in the musculature, causing cysticerci. As in *cysticercus bovis*, each cyst contains the infective cephalic end of the worm, which, when ingested by the definitive host (human), excysts and attaches itself to the small intestine, where the worm develops to the adult form. Symptoms are abdominal pain, diarrhea, and passage of segments through the anus.

Humans may also develop cysticercosis when they ingest the eggs of *Taenia solium* along with raw vegetables or drinking water. The larval forms migrate to and encyst in muscles, brain, and other tissues. Auto-infection may possibly cause cysticercosis in humans when reverse peristalsis carries eggs up to the stomach and duodenum, enabling hatching and migration of larval forms. The larvae in the cysticerci die off within a year and become calcified. Symptoms produced are related to the site of the lesion; for example, Jacksonian epilepsy can occur when the cysts form in the brain.

Prevention of transmission is achieved by (1) eating only well-cooked pork; (2) inspection of pig carcasses in the slaughterhouse to eliminate any that

show signs of cysticercosis infection; and (3) safe disposal of human feces to interrupt the transmission to pigs.

Hymenolepiasis

Hymenolepiasis is cosmopolitan in distribution. Children in institutions and the immunodeficient or malnourished are particularly prone to it. The dwarf tapeworms of *Hymenolepis* species do not require an intermediate host. The adult worm lives in the small intestine, passing its eggs in the feces. When these eggs are ingested by other humans along with contaminated food or water, infection occurs. Autoinfection also takes place when eggs hatch out in a patient's gut; they develop into adults in 2 weeks.

Hydatidosis (Echinococcosis)

Hydatid disease occurs in the sheep- and cattle-rearing areas of the world, mainly those in South America, Kenya, Vietnam, and China. The adult *Echinococcus granulosus* is small, less than 1 centimeter in length, and inhabits the small intestine of carnivorous animals like dogs – the definitive hosts. The terminal segments of the worm, which contains the eggs, are passed in the dog's feces and can survive for months on pastures. When eaten by the intermediate host (humans, cattle, sheep, goats), the eggs hatch and the larval form penetrates the intestinal wall to be carried by the bloodstream to various parts of the body where hydatid cysts develop. The sites affected are often the liver, lung, brain, or bone, with the slow-growing cyst consisting of a laminated cyst wall filled with fluid and containing many infective forms (hydatid sand). When the tissue form is eaten raw, as happens when dogs are fed infected offal in farming regions, adult worms develop in the dog to repeat the cycle.

Symptoms of a slow-growing, space-occupying cyst are related to the site of the lesion. Treatment often requires careful surgical removal of the cyst. Prevention involves personal hygiene to avoid the ingestion of eggs from dogs, with regular worming of dogs to eliminate the tapeworm, and sanitary disposal of infected offal and viscera of slaughtered animals.

Diphyllobothriasis

Diphyllobothriasis occurs in temperate and subarctic countries where fish is eaten raw, as in Finland, the former Soviet Union, Japan, Canada, the United States, and Chile. The adult *Diphyllobothrium latum* (fish tapeworm) measures 3 to 10 meters long and remains attached by the head end to the mucosa of the small intestine. Eggs are discharged on their own (not within the segments as in taeniasis), pass out in feces, and hatch when they reach fresh water. The parasite now waits to be ingested by a freshwater microcrustacean (cyclops), in which it develops further. If this tiny freshwater cyclops is ingested by a freshwater fish, the larva emerges and penetrates the intestinal

wall, developing into its infective form in the muscles of the fish. People ingesting infected raw fish acquire the disease, the symptoms of which include diarrhea, vomiting, fatigue, abdominal pain, dizziness, and megaloblastic anemia. Diagnosis is based on looking for eggs in stools. Prevention means avoiding raw fish, for the infective larvae can be killed by cooking, freezing, or thorough pickling procedures. Human sewage should be treated before discharge into water.

Nematodes (Roundworms)

Ascariasis

Ascariasis is a common disease in areas with poor sanitation, especially in the tropics. Infection with *Ascaris lumbricoides* (roundworms) is achieved by ingesting embryonated eggs on food, like salad vegetables. The larvae hatch, penetrate the wall of the duodenum, and travel in the bloodstream to the lungs, where they develop further. They then penetrate the alveoli and come up the trachea to be swallowed down into the intestines, where they settle in to mature in the ileum. Adults live in the lumen of the gut for up to 2 years, and their eggs are passed in the feces. The eggs are thick-walled and able to resist drying and the standard procedure for sewage treatment. They are, however, killed by heat over 60° C. These eggs require a further period of 10 to 50 days to develop in warm moist soil before they become infective. They may be consumed on raw salad vegetables or unpeeled fruit.

The majority of cases remain symptomless although respiratory symptoms occur when the larvae are traversing the lungs (cough, chest pain, pneumonitis with eosinophilia). When a heavy worm load is present in the intestine or when allergic reactions to the worm occur, symptoms of digestive disorder or intestinal obstruction may be observed. Diagnosis may be made by demonstrating eggs in feces. Prevention is accomplished by thorough washing of all salad vegetables, improvement of sanitation, safe disposal of sewage, and treatment of existing cases.

Trichuriasis

The adult nematode *Trichuris trichiura* (whipworm or thenoworm) has a narrow anterior whiplike portion and a broader posterior end. The whole worm is 40 centimeters long and lives in the colon with the anterior whiplike end embedded in the mucosa. Eggs are passed in feces, develop in warm moist soil, and are ingested by another human host. Larvae hatch to penetrate the wall of the large intestine where they develop into the adult form. They then return to the lumen of the colon, embed their whiplike ends by tunneling through mucosa, and repeat the cycle.

In children, following heavy infestation, the disease may present as retarded growth and malnutrition. Sometimes a prolapsed rectum may occur. Prevention measures are the same as those for ascariasis.

Enterobiasis

Enterobiasis is a common disease in most parts of the world, including temperate countries. The adult *Enterobius vermicularis* (pinworm) is about 11 millimeters long and lives in the lumen of the large bowel. Females lay their eggs on the skin around the anus, which causes intense itching. When scratched, the eggs are transferred on fingers or through dust to food. After they are ingested, larvae emerge to develop into adults in the bowel. The disease is common among children, can affect whole families, and is seen in institutions.

Diagnosis involves the demonstration of eggs in skin around the anus, whereas prevention involves hand washing and treatment of whole families or institutions. The cleaning of bedclothes and rooms to remove eggs is also essential.

Trichinosis

Infection with *Trichinella spiralis* in animals is worldwide, but human infection is mainly seen in Eastern Europe, the Arctic regions, South America, Asia, and East Africa. The domestic pig is the main source of infection, and the infective larval form is found encysted in the muscles. When eaten in raw or undercooked pork, it develops into the adult worm, which lives embedded in the mucosa of the small intestine. The adult is viviparous, producing many live larvae, which penetrate the intestinal wall to be carried in the blood to muscles, where they coil up forming cysts.

Symptoms include diarrhea, abdominal pain, and pain in the muscles. Heavy infection causes severe illness and may be fatal due to the involvement of the nervous or cardiac systems.

Diagnosis is confirmed by looking for larvae in muscle biopsies, and prevention is accomplished by cooking meat at over 60° C or deep freezing it to kill the larva. In addition, improving their feed prevents pigs from contracting the disease.

Sujatha Panikker

Bibliography

- Adak, G. K., E. O. Caul, and J. M. Cowden. 1991. Infection with small round structured viruses, England and Wales 1981–1990. *Communicable Disease Report of Central Public Health Laboratory Services* 1: 13, 141–4.
- Dunbar, E. M. 1990. Botulism. *Journal of Infection* 20: 1–3.
- Hughes, J. M. 1985. Food poisoning. In *Principles and practice of infectious diseases*. Second edition, ed. G. L. Mandel, R. G. Douglas, and J. E. Bennett, 680–91. New York.
- Muller, R., and J. R. Baker. 1990. *Medical parasitology*. London.
- Parker, M. T., and L. H. Collier. 1990. *Topley and Wilson's principles of bacteriology, virology and immunity*. 5 vols, 1: 243–308; 3: 49–54, 315–29, 423–572; 4: 323–413. London.
- Waites, W. M., and J. P. Arbutnott. 1990. Foodborne illness: An overview. *Lancet* 336: 722–5.

IV.E.5 Food Sensitivities: Allergies and Intolerances

Foods and beverages contain nutrients that are essential to human life, but they also contain elements that, for some individuals, may be harmful to health or even life-threatening. Foods and beverages may cause adverse reactions when ingested, but humans also can have adverse reactions to foods through inhalation (Edwards, McConnochie, and Davies 1985; Kemp, Van Asperen, and Douglas 1988), skin contact (Mathias 1983), or injection (Flu Shots and Egg Allergy 1992; Schwartz 1992).

Sensitivity to foods may be caused by immunologic abnormalities or by other mechanisms, such as host enzyme deficiency. Food sensitivities caused by immunologic abnormalities are commonly referred to as food allergies, whereas food sensitivities caused by nonimmunologic mechanisms are referred to as food intolerances (Anderson 1990; Beaudette 1991). This chapter reviews major considerations with regard to both classifications of food sensitivities. It should also be noted that a disorder called “pseudo-food allergy syndrome” exists. This syndrome is a psychological disorder in which the sufferer believes in the existence of a food allergy that cannot be confirmed by clinical testing (Pearson and Rix 1983, Pearson 1988).

Food Allergies

Allergic, or immunologically based, reactions to food most commonly involve a Type I or IgE mediated reaction (Anderson, 1990) as classified by R. R. A. Coombs and P. G. H. Gell (1975). The stages of the Type I reaction are illustrated in Table IV.E.5.1. It has been theorized that IgE mediated reactions to food may be related to a reduction in the need to produce IgE in response to intestinal parasitic infection. This theory, known as the “mast cell saturation hypothesis,” speculates that when the majority of IgE antibodies are bound to intestinal parasites, fewer antibodies are available to bind with food antigens. Production of high levels of IgE to fight parasitism would be protective, whereas production of high levels of IgE specific for food antigens is counterproductive to health (Godfrey 1975; Merrett, Merrett, and Cookson 1976; Lieberman and Barnes 1990).

In food allergy, the Type I reaction involves production of IgE antibodies that are tailored to food allergens. Food allergens are most commonly proteins but also may be glycoproteins or glycolipids (Bindels and Verwimp 1990). In order for a food allergen to stimulate IgE production, it must be absorbed and be of the appropriate size and shape to bridge IgE antibodies on the mast cell surface (Perkin 1990; Taylor 1990). Although most individuals are able to digest food proteins without eliciting an allergic response, individu-

Table IV.E.5.1. *The Type I allergic reaction*

Stage	Description
1st	Familial predisposition: To make IgE allergen-specific antibodies-Atopy.
2d	Sensitization and fixation: Specific allergen exposure, sensitization to form IgE allergen-specific antibodies followed by mast cell (or basophil effector cells) fixation.
3d	An-Ab interaction, mediator release/formation: reexposure to specific allergen, allergen +IgE antibody interaction followed by both preformed chemical mediator release from mast cells (or basophils) or activation of other mediator pathways outside the effector cells.
4th	Immune inflammation: Influx of leukocytes and monocytes. Release of other chemical mediators from all types of cells, plus tissue reactive substances from eosinophils. Release of cytokine histamine releasing factor (HRF) to promote nonspecific reactivity.
5th	Clinical reactions: (1) Early phase-within minutes to hours an allergic reaction occurs, including vasodilation, smooth muscle contraction, and hyperglandulin secretion. (2) Late phase - in one-third to one-half of all cases, in 6 to 12 hours a late inflammatory reaction occurs that is characterized by eosinophil (and other leukocytes and monocytes) infiltration - tissue damage and other mediator release (from all cells). (3) Prolonged phase - clinical hyperreactivity induced by nonspecific chemical mediator release, following nonallergic stimuli and modified by HRF.

Source: Perkin 1990: 4.

als who are atopic seem to absorb food antigens in a form that causes their cells to produce IgE. When food allergens are of the appropriate size, they are capable of binding with IgE and triggering the release of chemicals that cause the clinical manifestations of food allergy. Chemicals commonly implicated in the allergic process include histamine, leukotrienes and cytokines (Metcalf, 1991).

In some individuals, exercise may play a role in stimulating the food allergy response (Sheffer et al. 1983). Food-dependent anaphylaxis related to exercise is thought to be associated with chemical mediator release from mast cells in conjunction with an abnormal functioning autonomic nervous system, specifically increased parasympathetic activity and decreased sympathetic activity (Fukutomi et al. 1992). Abnormalities of IgE synthesis associated with other disorders may result in the secondary appearance of food allergy. D. S. Mazza, M. O'Sullivan, and M. H. Grieco (1991) have published a case report of food allergy that developed secondary to infection with the human immunodeficiency virus.

There are types of food-related immunologic reactions that are not believed to be Type I. Heiner's syndrome (a bovine milk-associated respiratory disorder) and celiac sprue are two examples of adverse food reactions believed to be associated with other types of immunologic reactions (Anderson 1990).

Food allergy is more common in children than adults although food allergy can develop at any age. One study that surveyed physicians estimated that allergy prevalence was 7 percent for adults and 13 percent for children (Anderson 1991). Another oft-cited study estimates the prevalence of allergy in children to be from 0.3 to 7.5 percent, with adults having a lower prevalence (Buckley and Metcalfe 1982). One theory used to explain the relatively higher level of allergy prevalence in children relates to infancy-associated intestinal immaturity, which, some believe, results in greater passage of larger proteins across the gut with resultant stimulation of IgE production (Walker 1975; Udall et al. 1981). But the primacy of a more permeable intestine in terms of allergy causation has recently been questioned by T. Jalonon (1991), who has postulated that greater intestinal permeability is a secondary phenomena for food allergy rather than an initiating event.

Infants may become sensitized to food allergens in utero since IgE can be synthesized before birth and large food proteins can cross the placenta. It should be noted, however, that although prenatal sensitization occurs, it is a relatively rare event (Strimas and Chi 1988). P. G. Calkhoven, M. Aalbers, V. L. Koshte, and colleagues (1991) have postulated the existence of a "high food responder phenotype" based on their findings that certain children seem to demonstrate allergic responses to a wide variety of foods. High cord IgE levels or high levels of serum IgE measured seven days after birth are considered to be predictive of subsequent food allergy development in an infant (Chandra, Puri, and Cheema 1985; Strimas and Chi 1988; Ruiz et al. 1991). A low level of CD8+ suppressor cells in the neonatal period is also predictive of the potential for allergy development (Chandra and Prasad 1991).

Breast feeding is often cited as being allergy protective, but this is an extremely controversial subject (Kramer 1988; Perkin 1990). Certainly there are physiological reasons to suggest that breast feeding may be protective against food allergies or at least may delay food allergy development. These reasons include (1) the relative lack of exposure to food antigens associated with breast feeding; (2) the presence of secretory IgA in breast milk, which is believed to aid in reducing the intestinal entry of large food protein molecules; (3) the stimulation of infantile IgA production by breast milk; (4) the opportunity for less intestinal antigen uptake because of the decreased incidence of gastrointestinal infections associated with breast feeding; (5) the presence in breast milk of anti-inflammatory properties, such as histaminase and arylsulfatase, and (6) the provision of IgG antibodies in breast milk that are targeted at food antigens (Pittard and Bill 1979; Atherton 1983; Businco et al. 1983; Ogra et al. 1984; Goldman et al. 1986; Michael et al. 1986). Low breast milk IgA levels have been associated with an increased likelihood of developing infantile cow's milk allergy (Savilahti et al. 1991).

Clinical studies that have examined breast feeding in relation to food allergy have not demonstrated a clear-cut protective or delaying effect (Kramer 1988; Perkin 1990), perhaps because of numerous methodological problems involved in this type of research. If breast feeding is protective, it is generally believed that exclusive breast feeding (no solid foods and no formula) should be practiced for at least six months (Bahna 1991; Kajosaari 1991). Since this is seldom done in the United States, the measurement of any allergy-prevention benefits of breast feeding is difficult.

Food allergies may disappear over time. Young age at food allergy onset coupled with a mild clinical reaction may mean that an individual will ultimately develop tolerance for an offending food. A more severe clinical reaction profile and older age of food allergy onset are associated with greater likelihood of allergy persistence. Allergies to certain foods, such as peanuts or crustacea, also tend to be allergies that persist through time (Anderson 1991). Food allergy developing in adulthood may be related to occupational exposure to food antigens (Anderson 1991; Metcalfe 1991). Baker's asthma, caused by allergic reactions to several wheat proteins, is one of the most studied occupational allergies (Prichard et al. 1985; Anderson 1991).

Certain foods have been commonly identified as implicated in the causation of food allergies. Among these are cow's milk, soybeans, wheat, eggs, nuts, and seeds (including cottonseed, which can be highly allergenic), crustacea (shrimp, lobster, crawfish), and fish. Also implicated are noncitrus fruits and vegetables (for example, tomato, celery, watermelon, pear, cherry, apple), citrus fruits, and spices (Perkin 1990). Table IV.E.5.2 outlines currently identified food allergens in some of the major food categories linked to allergy (Perkin 1990). Food types linked to allergy reflect cultural dietary patterns (Esteban 1992; Walker 1992).

The literature continues to identify new food categories that may cause allergic reactions in some individuals. Foods recently identified include squid (Carrillo et al. 1992), grand keyhole limpet and abalone (Morikawa et al. 1990), annatto dye (Nish et al. 1991), jícama (Fine 1991), and Swiss chard (de la Hoz et al. 1991). Progress has also been made in terms of identifying other factors, such as mites, that may be related to food allergies. For example, A. Armentia, J. Tapias, D. Barber, and colleagues (1992) have demonstrated that an allergic reaction may be precipitated by exposure to the wheat flour storage mite *Lepidoglyphus destructor*. Foods or beverages used as folk remedies may also elicit allergic reactions. J. Subiza, J. L. Subiza, M. Alonzo, and co-workers (1990) have described allergic responses that were initiated by washing the eyes with chamomile tea. Even human breast milk may contain allergenic food proteins (Gerrard and Shenassa 1983; Kilshaw and Cant 1984; Gerrard and Perelmutter 1986).

Table IV.E.5.2. Major food allergens

Food	Identified allergens
Cow's milk	Beta-lactoglobulin, alpha-lactalbumin, casein
Soybeans	Globulin 2S, Globulin 7S, Globulin 11S, hemagglutinin
Shrimp	Sa-I, Sa-II, antigen I, antigen II
Fish	Allergen M, protamine sulfate
Egg white	Ovalalbumin, ovomucoid
Egg yolk	Apovitellenin I, Apovitellenin VI

The determination of the extent to which foods related to one another in botanically defined plant families trigger allergies has long been of interest. When such a reaction occurs, it is called cross-reactivity. Legumes (Barnett, Bonham, and Howden 1987) and crustacea (Sachs and O'Connell 1988) are often studied in this regard. Cross-reaction between environmental and food allergens has also been reported. Birch pollen has been noted to be cross-reactive with the carrot and apple (Ortolani et al. 1988), and banana and latex have recently been reported as cross-reactive (M'Raihi et al. 1991). The phenomenon of cross-reactivity is one that is currently ill defined and is a subject of debate in terms of its clinical relevance.

Clinical signs and symptoms of food allergy are numerous, and diagnosis of food allergy is a complex process. Clinical manifestations may occur in selected body systems, such as the respiratory system or the skin, or may be a more generalized phenomenon, such as anaphylaxis. There are also specific syndrome symptom complexes that have been linked to food allergy.

Skin symptoms of food allergy may include hives, angioedema, eczema, and itching and/or redness (Anderson 1991). Eczema is particularly common in children and usually occurs on the elbows, knees, and perhaps the face (Burks 1992). Respiratory symptoms may include rhinitis, wheezing, and asthma.

Certain food allergies also produce gastrointestinal symptoms, such as nausea, vomiting, and diarrhea (Anderson 1991; Beaudette 1991; Metcalfe 1991). Cow's milk allergy is particularly noted for its association with gastrointestinal symptoms. It may also be associated with low-grade gastrointestinal bleeding and, thus, may be a cause of anemia (Wilson and Hamburger 1988).

Anaphylaxis is a systemic manifestation of food allergy. It is also potentially deadly, although it may be mild and simply characterized by simultaneous allergic manifestations in several organ systems. But the term is commonly used to refer to the more serious form of multiorgan system reaction that includes the development of cardiovascular shock, severe respiratory distress, and even death (Anderson 1990). Symptoms of classic food allergy generally appear within one to two hours after ingestion of the offending food, although in some cases, symptoms can appear

as much as 12 hours following consumption (Anderson 1990).

Several clinical syndromes have also been associated with food allergy. These include oral allergy syndrome, eosinophilic gastroenteritis, food protein-induced enterocolitis syndromes, and hypersensitive furrowed mouth (Anderson 1991; Metcalfe 1991).

Oral allergy syndrome, as described by C. Ortolani, M. Spano, E. Pastorello, and colleagues (1988), is initially manifested by irritation and swelling of the lips, which is then followed by hives, respiratory problems, and, in some cases, the development of anaphylactic shock. The oral allergy syndrome is associated with allergies to certain fruits and vegetables, notably celery (Ortolani et al. 1988; Pauli et al. 1988). J. A. Anderson (1991) has estimated that 0.1 to 0.2 percent of the general population may experience oral allergy syndrome. The syndrome is more frequently seen in persons who are sensitive to ragweed pollen.

Eosinophilic gastroenteritis associated with food allergy is characterized clinically by gastrointestinal disturbance, elevated levels of eosinophils, the presence of eosinophils in the stomach and intestine, and elevated levels of IgE (Anderson 1991; Metcalfe 1991). The syndrome of food protein-induced enterocolitis is seen in infants and is characterized by diarrhea, malabsorption, and high levels of eosinophils (Metcalfe 1991).

Hypersensitive furrowed mouth syndrome is characterized by mouth swelling and the development of cracks and furrows in the mouth. Anderson (1991), who described this problem, has associated it with consumption of large amounts of foods rich in protein.

Diagnosis of food allergy can be complex. It involves analysis of the clinical history, epicutaneous skin testing, and/or the use of *in vitro* assays, such as the radioallergosorbent test (RAST) or the enzyme-linked immunosorbent assay (ELISA) (Anderson 1990).

Elimination diets and the analysis of diet-symptom diaries may also be used in the diagnostic effort (Olejer 1990; Burks and Sampson 1992). The "gold" standard for diagnosis of food allergy is the double blind placebo-controlled food challenge (Olejer 1990). In children under the age of three years, skin testing is not a reliable indicator of food allergy, and single blind food challenge is more commonly employed to determine foods to which a child has an adverse reaction (Olejer 1990; Beaudette 1991). Diagnostic procedures used for research purposes are being studied for potential clinical application. These include basophil histamine release assay and assay of intestinal mast cell histamine release (Burks and Sampson 1992).

Once a food allergy is diagnosed, treatment primarily involves instructing the patient to eliminate the offending substance from the diet. This requires that the allergic individual pay careful attention to food

preparation techniques and food labeling. Unfortunately, food labels are not always helpful and can even be misleading. A label of kosher-parve, for example, generally denotes a milk-free product, but a recent report indicated that a dessert labeled as kosher-parve was found to contain milk protein (Jones, Squillace, and Yunginger 1992). In this instance, the milk proteins were present in the product as the result of a faulty production process that permitted milk contamination from previous use of the equipment.

In some cases, individuals may not understand the wording on food labels. R. N. Hamburger (1992) noted that one of his patients did not recognize that the term "calcium caseinate" meant the presence of milk protein. Eating outside of the home also presents special problems. The National Restaurant Association has recently launched an allergy awareness program and is attempting to make restaurant owners more aware of the needs of patrons with food allergies (Restaurant Food Allergy Awareness 1992).

Special attention must be paid to the nutritional adequacy of diets when foods are eliminated. Nutritional deficits, such as insufficiencies of calories, protein, and calcium, have been noted in children following restricted diets for food allergies (Lloyd-Still 1979; Sinatra and Merritt 1981; David, Waddington, and Stanton 1984).

For limited periods of time, elemental diet formulas may be useful for managing food allergies. These formulas contain protein that has been extensively hydrolyzed. However, the relatively poor taste and high cost of these products generally make them a poor long-term solution for food allergies. They can, however, help symptoms resolve and can bring relief prior to the development of a restricted dietary program of traditional foods.

Formula-fed infants who are allergic to cow's milk protein are placed on alternative formulas. These may include a soy-based formula or one of casein hydrolysate (Olejer 1990). The use of soy formulas, however, to treat infants with cow's milk allergy is controversial, and whey hydrolysate formulas are currently not recommended as treatment for such infants (American Academy of Pediatrics 1983; Businco et al. 1992).

Hypoallergenic formulas are defined as those that can be tolerated by children with cow's milk allergy, such that 90 percent will experience no symptoms at a 95 percent level of confidence (Sampson 1992). Casein hydrolysate formulas are currently regarded as hypoallergenic (Oldaeus et al. 1992; Rugo, Wahl, and Wahn 1992). Breast-fed infants who are allergic to food proteins transferred through breast milk may be treated by the mother's restriction of the offending foods from her diet (Perkin 1990).

Traditionally, pharmacological approaches have had limited application in the treatment of food allergies. Epinephrine is used as an antianaphylactic agent, and antihistamines and corticosteroids are employed

to alleviate allergy symptoms (Doering 1990). Oral cromolyn is effective in some cases for treating food allergy or other food-related problems, such as migraine (Doering 1990; Knottnerus and Pelikan 1993). Injection treatment with peanut extract is currently being investigated (Oppenheimer et al. 1992), as are such drugs as loratadine (Molkhou and Czarlewski 1993) and pancreatic enzyme supplements (Bahna and McCann 1993).

Research efforts are attempting to find ways that food allergies can be prevented or delayed. One thrust of the research is to examine infant formulas for those types that can delay or prevent allergies. Although not currently recommended for cow's milk allergy treatment, whey hydrolysates have been demonstrated by some studies to be protective (Chandra and Prasad 1991; Vandenplas et al. 1992); other studies have found that casein hydrolysate formula can decrease the occurrence of atopic dermatitis (Bahna 1991) and eczema (Mallet and Henocq 1992).

The majority of food allergy prevention focuses on alterations of both maternal and infant diet. As early as 1983, Hamburger, S. Heller, M. H. Mellon, and colleagues began testing the efficacy of a maternal elimination diet (no eggs, peanuts, and milk during the last trimester of pregnancy and during lactation) in conjunction with modifications in the infant's diet, such as delayed introduction of solids and breast feeding supplemented only by casein hydrolysate formula. But this type of regimen has yet to prove successful even though it may help to delay allergy appearance (Zeiger et al. 1986; Zeiger et al. 1989). Preventive strategies, such as the restrictive regimen just described, are used with infants who are considered at high allergy risk as determined by cord blood IgE levels (greater than 2 micrograms per liter) and a positive family history of allergy (Beaudette 1991). M. Kajosaari (1991) has demonstrated that delay of solid feeding until the age of six months can help prevent food allergy, and such a delay is currently standard dietary advice.

Food Intolerances

Some individuals experience adverse reactions to foods that cannot be explained by an immunologic mechanism. Several substances seem to elicit clinical problems through these alternative pathways that include enzyme deficiencies and pharmacological mechanisms (Schwartz 1992). Examples of food components associated with intolerance include lactose, histamine, gliadin, aspartame, sulfites, tartrazine, and monosodium glutamate. Some of these components are naturally present in foods, whereas others occur as food additives.

Lactose (or milk sugar) consumption causes problems because of inadequate amounts of intestinal lactase activity (MacDonald 1988). Symptoms of lactose intolerance include diarrhea, pain, and abdominal

bloating (Scrimshaw and Murray 1988). The most common type of lactose intolerance is termed primary lactase deficiency. It occurs sometime following weaning and is found in varying degrees from mild to severe. It is a common condition among the majority of the world's populations (Savaiano and Kotz 1988). Lactose intolerance can also occur as a secondary event subsequent to intestinal damage (Penny and Brown 1992).

Histamine, a preformed chemical mediator for allergic reactions, is naturally present in some foods and wines (Malone and Metcalfe 1986; Anderson 1990). Most of the time, histamine in foods does not cause problems because it is quickly broken down by enzymes in the digestive process. Some individuals do experience symptoms of histamine sensitivity, which include headache and reddening of various body parts, such as the eyes, face, and hands. Histamine sensitivity problems may also be seen when histidine is degraded to become histamine in fish and cheeses (Burnett 1990). Scombroid fish poisoning is the name given to the clinical complex (headache, flushing, and neck pain) caused by eating dark fish meat in which histidine has broken down to histamine because of improper storage and high temperatures. This syndrome can be treated with histamine antagonist drugs (Morrow et al. 1991).

Gliadin, a component of gluten, is a natural food substance that can cause adverse reactions in some individuals. It occurs in grain products such as wheat, rye, and barley. Celiac sprue, an intestinal disorder, and dermatitis herpetiformis, a skin disorder, are the clinical manifestations of intolerance to gliadin. A transient, rare form of gluten hypersensitivity, similar to celiac sprue, has also been reported in very young children (Iacono et al. 1991).

Celiac sprue, the intestinal form of gliadin intolerance, is characterized by iron deficiency, weight loss, diarrhea, malabsorption, abdominal distension, and altered mental capabilities. Persons suffering chronically from celiac sprue appear to be at increased risk for the development of cancer (Holmes et al. 1989) and osteoporosis later in life (Mora et al. 1993). Recent work has focused on identifying and describing peptides of gliadin that cause the damage in celiac sprue (Cornell, Weiser, and Belitz 1992). The disease may have an immunologic basis, but this has not been fully described; and for the moment at least, celiac sprue is classified as a food intolerance rather than a food allergy.

Dermatitis herpetiformis, the skin form of gliadin intolerance, is characterized by a rash with itching and blistering. Granular deposits of the immunoglobulin IgA between skin layers are also characteristic of dermatitis herpetiformis (Beaudette 1991).

Aspartame is a nutritive artificial sweetener added to a variety of foods. It was introduced into the marketplace in 1981 (Tollefson and Barnard 1992). Adverse reactions reported in conjunction with aspartame ingestion include headache, memory loss,

depression, dizziness, changes in vision, and seizures (Bradstock et al. 1986; Roberts 1990). L. Tollefson and R. J. Barnard (1992) have indicated, however, that an analysis by the United States Food and Drug Administration does not support the assertion that aspartame causes seizures. The postulated mechanism for aspartame's link to headache is via increasing serum tyrosine levels (Schiffman et al. 1987). One study has cited an immune basis for aspartame sensitivity manifested by hives (Kulczycki 1986). Aspartame, however, has not been shown to degranulate basophils or mast cells (Garriga and Metcalfe 1988). Therefore, in general, adverse reactions in conjunction with aspartame tend to be reported under food intolerance because of an unknown mechanism or mechanisms. After failing to reproduce hypersensitivity reactions to aspartame, M. M. Garriga, C. Berkebile, and D. D. Metcalfe (1991) noted that aspartame should not induce an IgE mediated response because it is easily broken down in the brush border of the intestine.

Sulfites are another category of food additive associated with adverse reactions in selected individuals. About one percent of the general United States population is estimated to be sulfite sensitive (Folkenberg 1988); this figure rises to about 5 percent of the asthmatic population (Nagy et al. 1993). As in the case of aspartame, most sulfite-associated adverse reactions cannot be linked to an immune mechanism, although an immune abnormality may be responsible for a reaction in a small number of individuals. W. N. Sokol and I. B. Hydock (1990) demonstrated basophil histamine release and obtained a positive skin test in a patient who had symptoms of angioedema, nasal congestion, and hives. Examples of clinical problems seen in conjunction with sulfite sensitivity include low blood pressure, hives, angioedema, intestinal cramping and chest tightness (Perkin 1990).

Sulfite sensitivity seems to have different clinical manifestations with asthmatics, where it is manifested by bronchial dysfunction (spasm or constriction), dizziness, flushing, wheezing, and perhaps anaphylaxis (Perkin 1990). Recently, sodium bisulfite has also been linked to genetic damage in lymphocytes (Meng and Zhang 1992). One postulated mechanism for sulfite sensitivity is the initiation of a cholinergic reflex by sulfur dioxide's action upon tracheobronchial receptors (Anibarro et al. 1992). With some individuals, it is believed that a deficiency of sulfate oxidase may be responsible (Simon 1987; Perkin 1990). Deaths and severe reactions related to sulfite sensitivity have resulted in strict labeling regulations when sulfites are present in foods in excess of 10 parts per million (ppm) (Schultz 1986). Experimentation is also under way to determine if the use of cyanocobalamin can prevent sulfite-induced respiratory distress (Anibarro et al. 1992).

Tartrazine, or FD&C Yellow 5, is another food additive cited as causing adverse reactions in susceptible individuals. Tartrazine is used to produce various food

colors including yellow, green, and maroon (Dong 1984; Schneider and Codispoti 1988), and more than half of the daily food dye consumption in the United States is in the form of tartrazine (Beaudette 1991). Clinical symptoms indicating an adverse reaction to this food additive include hives, asthma, angioedema, photosensitivity, eczema, purpura, and anaphylaxis (Michaelsson, Petterson, and Juhlin 1974; Desmond and Trautlein 1981; Pereyo 1987; Perkin 1990; Devlin and David 1992). J. Devlin and T. J. David (1992) recently indicated that they could confirm tartrazine intolerance in only 1 of 12 children whose parents reported that tartrazine consumption worsened their child's eczema. Although the mechanism of tartrazine sensitivity is unknown, the cause has been postulated as an excess of bradykinin production (Neuman et al. 1978). There is a link between aspirin sensitivity and tartrazine sensitivity with a cross-reactivity of 5 to 25 percent reported (Settipane and Pudupakkam 1975; Condemi 1981).

Monosodium glutamate (MSG) as the cause of adverse food reaction has been a source of great controversy. It is used as a food additive because of its ability to enhance flavors (Beaudette 1991). The most often cited potential adverse reaction to MSG is Chinese Restaurant Syndrome characterized by such symptoms as development of tears in the eyes, facial flushing, tightness and burning, nausea, sweating, headache, and vascular abnormalities (Ghadimi and Kumar 1972; Gann 1977; Goldberg 1982; Zautcke, Schwartz, and Mueller 1986).

Symptoms of this syndrome usually occur shortly after consumption of an MSG-containing food, and problems generally resolve in less than one hour (Beaudette 1991). Some scientists suggest that the role of MSG as the causative agent of the Chinese Restaurant Syndrome remains unproven because of a lack of a demonstrated dose-response effect (Beaudette 1991), and histamine has been proposed as an alternative culprit (Chin, Garriga, and Metcalfe 1989). In addition to Chinese Restaurant Syndrome, other potential adverse reactions to MSG consumption have been reported. These include headache, angioedema, and asthma (Diamond, Prager, and Freitag 1986; Allen, Delohery, and Baker 1987; Squire 1987). It has also been suggested that there may be a subset of the population for whom the potential excitotoxicity of glutamate could be a problem (Barinaga 1990).

Prevention and treatment of food intolerance involves avoidance of the foods and beverages that contain the offending substances. Careful reading of food labels is critical. In the case of lactose intolerance, some individuals can consume small amounts of milk or dairy products, such as yogurt or cheese, which have a low lactose content. In addition, some find it beneficial to employ special dairy products in which the lactose has been partially hydrolyzed or broken down or to use tablets that serve to break down lactose in foods and beverages. Celiac sprue

and dermatitis herpetiformis are treated with a gluten-restricted, gliadin-free diet (Beaudette 1991), sometimes referred to as a gluten-free diet. The gluten-free diet has recently been shown to relieve the primary symptoms of these conditions, and also to aid in protecting against bone loss (Mora et al. 1993) and cancer (Holmes et al. 1989). Some have advocated the use of a low gluten diet for the treatment of celiac sprue (Kumar et al. 1985; Montgomery et al. 1988), but at present this is in the exploration stage as a potential treatment alternative.

Summary

Foods and beverages may cause clinical problems through immunologic or physiological mechanisms. When immunologic mechanisms are confirmed, the condition is called a food allergy. The term food intolerance is used to denote clinical problems associated with food when an immunologic mechanism cannot be confirmed.

Common food allergens in the United States include cow's milk, soybeans, wheat, eggs, nuts, seeds, and crustacea. Food allergies are more common in children than adults. Most food allergies are of the Type I IgE mediated variety. Avoidance of the foods that cause problems is the mainstay of current prevention and treatment efforts.

Food intolerance may be associated with natural food components or with food additives. Natural food components that cause clinical problems include lactose (milk sugar), histamine, and gliadin. Food additives linked to intolerance include aspartame, sulfites, monosodium glutamate, and tartrazine. Here again, dietary avoidance is the key to preventing and treating food intolerances, although pharmacological agents may be useful in treating symptoms associated with food allergies or intolerances. Several new pharmacological approaches are now being investigated.

Research in the area of food intolerances continues to look at immune mechanisms (particularly those in addition to Type I) and other physiological mechanisms that cause clinical problems related to exposure to foods and beverages. Research efforts are also underway to characterize the specific components that cause food allergies and to find new ways of treating them. In addition, testing of potential preventive or delaying approaches will continue to be a research focus in future years.

Judy Perkin

Bibliography

- Allen, D. H., J. Delohery, and G. Baker. 1987. Monosodium-L-glutamate-induced asthma. *Journal of Allergy and Clinical Immunology* 80: 530-7.
- American Academy of Pediatrics Committee on Nutrition. 1983. Soy protein formulas: Recommendations for use in infant feeding. *Pediatrics* 72: 359.
- Anderson, J. A. 1990. Food allergy or sensitivity terminology, physiologic bases and scope of the clinical problem. In *Food allergies and adverse reactions*, ed. J. E. Perkin, 1-13. Gaithersburg, Md.
1991. The clinical spectrum of food allergy in adults. *Clinical and Experimental Allergy* 21 (Suppl. 1): 305-15.
- Anibarro, B., T. Caballero, C. Garcia-Ara, et al. 1992. Asthma with sulfite intolerance in children: A blocking study with cyanocobalamin. *Journal of Allergy and Clinical Immunology* 90: 103-9.
- Armentia, A., J. Tapias, D. Barber, et al. 1992. Sensitization to the storage mite *Lepidoglyphus destructor* in wheat flour allergy. *Annals of Allergy* 68: 398-403.
- Atherton, D. J. 1983. Breastfeeding and atopic eczema. *British Medical Journal* 287: 775-6.
- Bahna, S. L. 1991. Breast milk and special formulas in prevention of milk allergy. *Advances in Experimental Medicine and Biology* 310: 445-51.
- Bahna, S. L., and M. L. McCann. 1993. Tolerance to multiple food allergens by pancreatic enzyme supplement. *Journal of Allergy and Clinical Immunology* 91: 151.
- Barinaga, M. 1990. MSG: Amino acids: How much excitement is too much? *Science* 247: 20-2.
- Barnett, D., B. Bonham, and M. E. H. Howden. 1987. Allergic cross-reactions among legume foods - an *in vitro* study. *Journal of Allergy and Clinical Immunology* 79: 433-8.
- Beaudette, T. 1991. *Adverse Reactions to Food*. Chicago.
- Bindels, J. G., and J. Verwimp. 1990. *Allergenic aspects of infant feeding - Nutricia Research Communications* 2. Zoetermeer, the Netherlands.
- Bradstock, M. K., M. K. Serdula, J. S. Marks, et al. 1986. Evaluation of reactions to food additives: The aspartame experience. *American Journal of Clinical Nutrition* 43: 464-9.
- Buckley, R. H., and D. Metcalfe. 1982. Food allergy. *Journal of the American Medical Association* 248: 2627-31.
- Burks, A. W. 1992. Food allergies and eczema. *Food Allergy News* 2: 1, 6.
- Burks, A. W., and H. A. Sampson. 1992. Diagnostic approaches to the patient with suspected food allergies. *Journal of Pediatrics* 121: S64-71.
- Burnett, J. W. 1990. Histamine poisoning. *Cutis* 46: 388.
- Businco, L., G. Bruno, P. G. Giampietro, and A. Cantani. 1992. Allergenicity and nutritional adequacy of soy protein formulas. *Journal of Pediatrics* 121: S21-8.
- Businco, L., F. Marchetti, G. Pellegrini, et al. 1983. Prevention of atopic disease in "at risk newborns" by prolonged breastfeeding. *Annals of Allergy* 51: 296-9.
- Calkhoven, P. G., M. Aalbers, V. L. Koshte, et al. 1991. Relationship between IgG 1 and IgG 4 antibodies to foods and the development of IgE antibodies to inhalant allergens. I. Establishment of a scoring system for the overall food responsiveness and its application to 213 unselected children. *Clinical and Experimental Allergy* 2: 91-8.
- Carrillo, T., R. Castillo, J. Caminero, et al. 1992. Squid hypersensitivity: A clinical and immunologic study. *Annals of Allergy* 68: 483-7.
- Chandra, R. K., and C. Prasad. 1991. Strategies for the prevention of food allergic associated atopic disease. *Advances in Experimental Medicine and Biology* 310: 391-6.
- Chandra, R. K., S. Puri, and P. S. Cheema. 1985. Predictive value of cord blood IgE in the development of atopic

- disease and role of breast-feeding in its prevention. *Clinical Allergy* 15: 517-22.
- Chin, K. W., M. M. Garriga, and D. D. Metcalfe. 1989. The histamine content of oriental foods. *Food Chemistry and Toxicology* 27: 283-7.
- Condemni, J. J. 1981. Aspirin and food dye reactions. *Bulletin of the New York Academy of Medicine* 57: 600-7.
- Coombs, R. R. A., and P. G. H. Gell. 1975. Classification of allergic reactions responsible for clinical hypersensitivity and disease. In *Clinical aspects of immunology*, ed. R. R. A. Coombs and P. G. H. Gell, 761-81. Oxford.
- Cornell, H., H. Weiser, and H-D. Belitz. 1992. Characterization of the gliadin-derived peptides which are biologically active in coeliac disease. *Clinica Chimica Acta* 213: 37-50.
- David, T. J., E. Waddington, and R. H. J. Stanton. 1984. Nutritional hazards of elimination diets in children with atopic eczema. *Archives of Diseases of Childhood* 59: 323-5.
- de la Hoz, B., M. Fernandez-Rivas, S. Quirce, et al. 1991. Swiss chard hypersensitivity: Clinical and immunologic study. *Annals of Allergy* 67: 487-92.
- Desmond, R. E., and J. J. Trautlein. 1981. Tartrazine (FD&C yellow #5) anaphylaxis: A case report. *Annals of Allergy* 46: 81-2.
- Devlin, J., and T. J. David. 1992. Tartrazine in atopic eczema. *Archives of Diseases of Childhood* 67: 709-11.
- Diamond, S., J. Prager, and F. G. Freitag. 1986. Diet and headache - is there a link? *Postgraduate Medicine* 1986 79: 279-86.
- Doering, P. 1990. Drug therapy of food allergies. In *Food allergies and adverse reactions*, ed. J. E. Perkin, 69-79. Gaithersburg, Md.
- Dong, F. M. 1984. *All about food allergy*. Philadelphia, Pa.
- Edwards, J. H., K. McConnochie, and B. H. Davies. 1985. Skin-test reactivity to egg protein - exposure by inhalation compared to ingestion. *Clinical Allergy* 15: 147-50.
- Esteban, M. M. 1992. Adverse food reactions in childhood: Concept, importance, and present problems. *Journal of Pediatrics* 121: S1-3.
- Fine, A. J. 1991. Hypersensitivity reaction to jicama (Pachyrhizus, yam bean). *Annals of Allergy* 66: 173-4.
- Flu Shots and Egg Allergy. 1992. *Food Allergy News* 2: 7.
- Folkenberg, J. 1988. Reporting reaction to additives. In *Safety first: Protecting America's food supply*, FDA Consumer Special Report 88-2224. Rockville, Md.
- Fukutomi, O., N. Kondo, H. Agata, et al. 1992. Abnormal responses of the autonomic nervous system in food-dependent exercise-induced anaphylaxis. *Annals of Allergy* 68: 438-45.
- Gann, D. 1977. Ventricular tachycardia in a patient with "Chinese restaurant syndrome." *Southern Medical Journal* 70: 879-81.
- Garriga, M. M., C. Berkebile, and D. D. Metcalfe. 1991. A combined single-blind, double-blind, placebo-controlled study to determine the reproducibility of hypersensitivity reactions to aspartame. *Journal of Allergy and Clinical Immunology* 87: 821-7.
- Garriga, M. M., and D. D. Metcalfe. 1988. Aspartame intolerance. *Annals of Allergy* 61 (Part II): 63-9.
- Gerrard, J. W., and L. Perelmutter. 1986. IgE-mediated allergy to peanut, cow's milk, and egg in children with special reference to maternal diet. *Annals of Allergy* 56: 351-4.
- Gerrard, J. W., and M. Shenassa. 1983. Sensitization to substance in breast milk: Recognition, management, and significance. *Annals of Allergy* 51: 300-2.
- Ghadimi, H., and S. Kumar. 1972. Current status of monosodium glutamate. *American Journal of Clinical Nutrition* 25: 643-6.
- Godfrey, R. C. 1975. Asthma and IgE levels in rural and urban communities of the Gambia. *Clinical Allergy* 5: 201-7.
- Goldberg, L. 1982. Supraventricular tachyarrhythmia in association with the Chinese restaurant syndrome. *Annals of Emergency Medicine* 11: 333.
- Goldman, A. S., L. W. Thorpe, R. M. Goldblum, and L. A. Hanson. 1986. Anti-inflammatory properties of human milk. *Acta Paediatrica Scandinavica* 75: 689-95.
- Hamburger, R. N. 1992. Misleading hazardous food labels. *Annals of Allergy* 68: 200-1.
- Hamburger, R. N., S. Heller, M. H. Mellon, et al. 1983. Current status of the clinical and immunological consequences of a prototype allergic disease prevention. *Annals of Allergy* 51: 281-90.
- Holmes, G. K. T., P. Prior, M. R. Lane, et al. 1989. Malignancy in celiac disease: Effect of a gluten-free diet. *Gut* 30: 333-8.
- Iacono, G., A. Nocerino, S. Guadalini, et al. 1991. Transient gluten hypersensitivity. *Journal of Pediatric Gastroenterology and Nutrition* 12: 400-3.
- Jalonen, T. 1991. Identical intestinal permeability changes in children with different clinical manifestations of cow's milk allergy. *Journal of Allergy and Clinical Immunology* 88: 737-42.
- Jones, R. T., D. L. Squillace, and J. W. Yunginger. 1992. Anaphylaxis in a milk-allergic child after ingestion of milk-contaminated kosher-pareve-labeled "dairy-free" dessert. *Annals of Allergy* 68: 223-7.
- Kajosaari, M. 1991. Atopy prophylaxis in high risk infants. *Advances in Experimental Medicine and Biology* 310: 453-8.
- Kemp, A. S., P. P. Van Asperen, and J. Douglas. 1988. Anaphylaxis caused by inhaled pavlova mix in egg-sensitive children. *Medical Journal of Australia* 149: 712-3.
- Kilshaw, P. J., and A. J. Cant. 1984. The passage of maternal dietary proteins into human breast milk. *International Archives of Allergy and Applied Immunology* 75: 8-15.
- Knottnerus, I., and Z. Pelikan. 1993. Protective effects of oral cromolyn (OSCG) on migraine due to the adverse reactions to foods. *Journal of Allergy and Clinical Immunology* 91: 150.
- Kramer, M. S. 1988. Does breastfeeding help protect against atopic disease? Biology, methodology, and a golden jubilee of controversy. *Journal of Pediatrics* 112: 181-90.
- Kulczycki, A., Jr. 1986. Aspartame-induced urticaria. *Annals of Internal Medicine* 104: 207-8.
- Kumar, P. J., G. Harris, J. Colyer, et al. 1985. Is a gluten-free diet necessary for the treatment of celiac disease? *Gastroenterology* 88: 1459.
- Lieberman, L. S., and K. C. Barnes. 1990. Food allergies and adverse food reactions: An anthropological perspective. In *Food allergies and adverse reactions*, ed. J. E. Perkin, 207-32. Gaithersburg, Md.
- Lloyd-Still, J. D. 1979. Chronic diarrhea of childhood and the misuse of elimination diets. *Journal of Pediatrics* 95: 10-13.
- MacDonald, I. 1988. Carbohydrates. In *Modern nutrition in health and disease*, ed. M. E. Shils and V. R. Young, 38-51. Philadelphia, Pa.
- Mallet, E., and A. Henocq. 1992. Long term prevention of allergic diseases by using a protein hydrolysate formula in at-risk infants. *Journal of Pediatrics* 121: S95-100.
- Malone, M. H., and D. D. Metcalfe. 1986. Histamine in foods: Its possible role in nonallergic adverse reactions to ingestants. *New England Regional Allergy Proceedings* 7: 241-5.
- Mathias, C. G. T. 1983. Contact urticaria from peanut butter. *Contact Dermatitis* 9: 66-8.

- Mazza, D. S., M. O'Sullivan, and M. H. Grieco. 1991. HIV-1 infection complicated by food allergy and allergic gastroenteritis: A case report. *Annals of Allergy* 66: 436–40.
- Meng, Z., and L. Zhang. 1992. Cytogenetic damage induced in human lymphocytes by sodium bisulfite. *Mutation Research* 298: 63–9.
- Merrett, T. G., J. Merrett, and J. B. Cookson. 1976. Allergy and parasites: The measurement of total and specific IgE levels in urban and rural communities in Rhodesia. *Clinical Allergy* 6: 131–4.
- Metcalfe, D. D. 1991. Immune mechanisms in food allergy. *Clinical and Experimental Allergy* 21 (Suppl. 1): 321–4.
- Michael, F. B., J. Bousquet, A. Dannaes, et al. 1986. Preventive measures in early childhood allergy. *Journal of Allergy and Clinical Immunology* 78: 1022–7.
- Michaelsson, G., F. Pettersson, and L. Juhlin. 1974. Purpura caused by food and drug additives. *Archives of Dermatology* 109: 49–52.
- Molkhou, P., and W. Czarlewski. 1993. Effect of loratadine on increased intestinal permeability in children with atopic dermatitis related to food allergy. *Journal of Allergy and Clinical Immunology* 91: 150.
- Montgomery, A. M. P., A. K. J. Goka, P. H. Kumar, et al. 1988. Low gluten diet in the treatment of adult celiac disease: Effect on jejunal morphology and serum anti-gluten antibodies. *Gut* 29: 1564–8.
- Mora, S., G. Weber, G. Barera, et al. 1993. Effect of gluten-free diet on bone mineral content in growing patients with celiac disease. *American Journal of Clinical Nutrition* 57: 224–8.
- Morikawa, A., M. Kato, K. Tokuyama, et al. 1990. Anaphylaxis to grand keyhole limpet (abalone-like shellfish) and abalone. *Annals of Allergy* 65: 415–7.
- Morrow, J. D., G. R. Margolies, J. Rowland, and L. J. Roberts. 1991. Evidence that histamine is the causative toxin of scombroid-fish poisoning. *New England Journal of Medicine* 324: 716–20.
- M'Raihi, L., D. Charpin, A. Pons, et al. 1991. Cross-reactivity between latex and banana. *Journal of Allergy and Clinical Immunology* 87: 129–30.
- Nagy, S. M., S. S. Teuber, P. Murphy, and S. M. Loscutoff. 1993. Adverse reactions to a salsa containing high levels of sulfites. *Journal of Allergy and Clinical Immunology* 91: 150.
- Neuman, I., R. Eliau, H. Nahum, et al. 1978. The danger of 'yellow dyes' (tartrazine) to allergic subjects. *Clinical Allergy* 8: 65–8.
- Nish, W. A., B. A. Whisman, D. W. Goetz, and D. A. Ramirez. 1991. Anaphylaxis to annatto dye: A case report. *Annals of Allergy* 66: 129–31.
- Ogra, P. L., R. C. Welliver, M. Riepenhoff-Talty, and H. S. Foden. 1984. Interaction of mucosal immune system and infections in infancy: Implications in allergy. *Annals of Allergy* 53: 523–34.
- Oldaus, G., C. K. Bradley, B. Bjorksten, and N. I. Kjellman. 1992. Allergenicity screening of "hypoallergenic" milk-based formulas. *Journal of Allergy and Clinical Immunology* 90: 133–5.
- Olejer, V. 1990. Diagnosis of food allergy. Part II. The dietitian's perspective. In *Food allergies and adverse food reactions*, ed. J. E. Perkin, 30–49. Gaithersburg, Md.
- Oppenheimer, J. J., H. S. Nelson, S. A. Bock, et al. 1992. Treatment of peanut allergy with rush immunotherapy. *Journal of Allergy and Clinical Immunology* 90: 256–62.
- Ortolani, C., M. Ispano, E. Pastorello, et al. 1988. The oral allergy syndrome. *Annals of Allergy* 61: 47–52.
- Pauli, G., J. C. Bessot, P. A. Braun et al. 1988. Celery allergy: Clinical and biological study of 20 cases. *Annals of Allergy* 60: 243–6.
- Pearson, D. J. 1988. Psychologic and somatic interrelationships in allergy and pseudoallergy. *Journal of Allergy and Clinical Immunology* 81: 351–60.
- Pearson, D. J., and K. J. B. Rix. 1983. Food allergy: How much is in the mind? A clinical and psychiatric study of suspected food hypersensitivity. *Lancet* 1: 259–61.
- Penny, M. E., and K. H. Brown. 1992. Lactose feeding during persistent diarrhoea. *Acta Paediatrica* Suppl. 381: 133–8.
- Pereyo, N. 1987. Tartrazine, the complement system and photosensitivity. *Journal of the American Academy of Dermatology* 17: 143.
- Perkin, J. E. 1990. Major food allergens and principles of dietary management. In *Food allergies and adverse reactions*, ed. J. E. Perkin, Gaithersburg, Md.
- Pittard, W. B., and K. Bill. 1979. Immunoregulation by breast milk cells. *Cellular Immunology* 42: 437–41.
- Prichard, M. G., G. Ryan, B. J. Walsh, and A. W. Musk. 1985. Skin test and RAST responses to wheat and common allergens and respiratory disease in bakers. *Clinical Allergy* 15: 203–10.
- Restaurant food allergy awareness program. 1992. *Food Allergy News* 2: 2.
- Roberts, H. J. 1990. Obstacles confronting consumer advocates: An overview of health-related issues. *Journal of the Florida Medical Association* 77: 761–5.
- Rugo, E., R. Wahl, and U. Wahn. 1992. How allergenic are hypoallergenic infant formulae? *Clinical and Experimental Allergy* 22: 635–9.
- Ruiz, R. G., D. Richards, D. M. Kemeny, and J. F. Price. 1991. Neonatal IgE: A poor screen for atopic disease. *Clinical and Experimental Allergy* 21: 467–72.
- Sachs, M. I., and E. J. O'Connell. 1988. Cross-reactivity of foods – mechanisms and clinical significance. *Annals of Allergy* 61: 36–40.
- Sampson, H. A. 1992. Adverse food reactions. *Journal of Pediatric Gastroenterology and Nutrition* 15: 319–20.
- Savaiano, D. A., and C. Kotz. 1988. Recent advances in the management of lactose intolerance. *Contemporary Nutrition* 13 (General Mills Nutrition Department Newsletter): 9, 10.
- Savilahti, E., V-M. Tainio, L. Salmenpera, et al. 1991. Levels of IgA and cow's milk antibodies in breast milk vs. the development of atopy in children. Low colostral IgA associated with cow's milk allergy. *Advances in Experimental Medicine and Biology* 30: 417–25.
- Schiffman, S. S., C. E. Buckley, H. A. Sampson, et al. 1987. Aspartame and susceptibility to headache. *New England Journal of Medicine* 317: 1181–5.
- Schneider, A. T., and A. J. Codispoti. 1988. Allergic reactions to food additives. In *Food allergy, a practical approach to diagnosis and management*, ed. L. T. Chiamonte, A. T. Schneider, and F. Lifshitz, 117–51. New York.
- Schultz, C. M. 1986. Sulfite sensitivity. *American Journal of Nursing* 86: 914.
- Schwartz, R. H. 1992. Allergy, intolerance, and other adverse reactions to foods. *Pediatric Annals* 21: 654–62, 665–74.
- Scrimshaw, N. S., and E. B. Murray. 1988. The acceptability of milk and milk products in populations with a high prevalence of lactose intolerance. *American Journal of Clinical Nutrition* 48 (Suppl): 1083–159.
- Settipane, G. A., and R. K. Pudupakkam. 1975. Aspirin intolerance, III. subtypes, familial occurrence and cross-reactivity with tartrazine. *Journal of Allergy and Clinical Immunology* 56: 215–21.
- Sheffer, A. L., N. A. Soter, E. R. McFadden, Jr., and K. F. Austen. 1983. Exercise-induced anaphylaxis: A distinct form of physical allergy. *Journal of Allergy and Clinical Immunology* 71: 311–16.

- Simon, R. A. 1987. Sulfite sensitivity. *Annals of Allergy* 59: 100-5.
- Sinatra, F. R., and R. J. Merritt. 1981. Iatrogenic kwashiorkor in infants. *American Journal of Diseases of Childhood* 135: 21-3.
- Sokol, W. N., and I. B. Hydock. 1990. Nasal congestion, urticaria, and angioedema caused by an Ig-E mediated reaction to sodium metabisulfite. *Annals of Allergy* 65: 233-7.
- Squire, E. N. 1987. Angio-oedema and monosodium glutamate. *Lancet* 1: 988.
- Strimas, J. H., and D. S. Chi. 1988. Significance of IgE level in amniotic fluid and cord blood for the prediction of allergy. *Annals of Allergy* 61: 133-6.
- Subiza, J., J. L. Subiza, M. Alonso, et al. 1990. Allergic conjunctivitis to chamomile tea. *Annals of Allergy* 65: 127-32.
- Taylor, S. L. 1990. Food allergies and related adverse reactions to foods: A food science perspective. In *Food allergies and adverse reactions*, ed. J. E. Perkin, 189-206. Gaithersburg, Md.
- Tollefson, L., and R. J. Barnard. 1992. An analysis of FDA passive surveillance reports of seizures associated with consumption of aspartame. *Journal of the American Dietetic Association* 92: 598-601.
- Udall, J. N., P. Colony, L. Fritze, et al. 1981. Development of gastrointestinal mucosal barrier, II. the effect of natural versus artificial feeding on intestinal permeability to macromolecules. *Pediatric Research* 15: 245-9.
- Uragoda, C. G. 1980. Histamine poisoning in tuberculosis patients after ingestion of tuna fish. *American Review of Respiratory Diseases* 121: 157-9.
- Vandenplas, Y., B. Hauser, C. Van den Borre, et al. 1992. Effect of a whey hydrolysate prophylaxis on atopic disease. *Annals of Allergy* 68: 419-24.
- Walker, W. A. 1975. Antigen absorption from the small intestine and gastrointestinal disease. *Pediatric Clinics of North America* 22: 731-46.
1992. Summary and future directions. *Journal of Pediatrics* 121: S4-6.
- Wilson, N. W., and R. N. Hamburger. 1988. Allergy to cow's milk in the first year of life and its prevention. *Annals of Allergy* 61: 323-7.
- Zautcke, J. L., J. A. Schwartz, and E. Mueller. 1986. Chinese restaurant syndrome: A review. *Annals of Emergency Medicine* 15: 1210-13.
- Zeiger, R. S., S. Heller, M. H. Mellon, et al. 1986. Effectiveness of dietary manipulation in the prevention of food allergy in infants. *Journal of Allergy and Clinical Immunology* 78: 224-38.
1989. Effect of combined maternal and infant food-allergen avoidance on development of atopy in early infancy: A randomized study. *Journal of Allergy and Clinical Immunology* 84: 72-89.

cal symptoms after milk consumption are described as lactose intolerant.

Lactose is a major constituent of the milk of all mammals except sea lions (Kretchmer 1993). Human milk contains the highest lactose concentration, about 7 percent; lactose levels in commonly milked animals, such as camels, goats, sheep, and cows, run between 4 and 5 percent. Adult animals, like most humans, lose the ability to digest lactose. This suggests that adult loss of lactase is a normal mammalian trait and that adult ability to split lactose is an "abnormal" evolutionary innovation.

Definition and History

Lactose malabsorption and intolerance must be distinguished clinically from allergy to milk proteins, which is a rare but serious genetic problem in infants. This essay focuses on primary adult onset lactase deficiency, but two other forms of the syndrome must be noted. Lactase deficiency may be secondary due to damage to the small intestine from heavy parasitic infections (especially the protozoan *Giardia lamblia*); to other severe intestinal infections; to AIDS; and to ionizing radiation, some drugs, and gastric surgery (Castiglia 1994; Tamm 1994). Total inability to synthesize lactase is another rare genetic disorder that was obviously lethal until modern times. If diagnosed promptly, such cases can now be managed with soy-based infant formulas.

Gastrointestinal distress in adults after milk consumption was described in ancient Greek and Roman texts, and there were isolated clinical reports in the late nineteenth and early twentieth centuries, but the problem was not widely studied until the development (in the 1960s) of new techniques to study enzymatic action in the intestine. Consequently, the high prevalence of diminished lactase activity in healthy adults was described only in the early 1960s, with especially important work done by A. Dahlqvist and his associates (Dahlqvist 1977). Worldwide surveys in the 1960s and 1970s showed that loss of lactase activity in adulthood is the common condition in humans and that terms like "lactase deficient" incorrectly imply that this is somehow abnormal (Flatz 1987).

Biology and Clinical Manifestations

Lactase (technically lactase-phlorizin hydrolase), is a protein produced in the cells of the epithelium of the small intestine. It is most concentrated in the mucosal cells of the brush border of the jejunum (Buller and Grand 1990). Production of lactase begins to decline in most children between the ages of 2 to 5, around the time of weaning. Most adults retain only about 10 percent of infant-level lactase activity. But Finnish children who become lactase deficient often do so as teenagers; the reasons for this late onset are unknown (Arola and Tamm 1994).

If lactose-intolerant people consume significant quantities of milk or other dairy products, unmetabo-

IV.E.6 🐾 Lactose Intolerance

Lactose is a disaccharide composed of linked molecules of the simple sugars glucose and galactose. Dietary lactose is obtained almost exclusively from milk. Infants and young children digest lactose with an enzyme, lactase, which splits the molecule into the two readily absorbable simple sugars. The majority of adults, however, have lost this ability and are lactose malabsorbers. Those malabsorbers who display clini-

lized lactose passes through the small intestine to the large intestine, where it is acted upon by the resident facultative bacterial flora. These bacteria split lactose into acetic, butyric, propionic, and other short-chain fatty acids, which can be absorbed by intestinal cells and used as metabolites. Among the by-products are carbon dioxide, hydrogen, and methane, which can cause a gassy, bloated, and/or nauseous feeling.

It is generally thought that the abundance of short-chain molecules increases osmotic pressure within the intestinal lumen, causing water to pass into the lumen, sometimes in amounts that produce diarrhea (Castiglia 1994), but this has recently been questioned by H. Arola and A. Tamm (1994). They suggest that bacteria that produce larger amounts of iso-fatty acids – those with branched carbon chains in contrast to the normal straight chain forms – may provide protection against diarrhea. Although the intestinal flora of an individual tend to remain relatively stable over time, commensal bacterial populations vary considerably among people. Those persons with large colonies of the types of bacteria that efficiently metabolize lactose and produce significant quantities of iso-fatty acids (for example, members of the genus *Bacteriodes*) would be less likely to display symptoms (Arola and Tamm 1994). They would be lactose malabsorbers but not necessarily lactose intolerant.

Genetics

The mechanisms controlling lactase production were disputed for many years. Some researchers, drawing on studies of gene regulation in bacteria, argued in the 1960s that lactase was a substrate-inducible enzyme; that is, that lactase production was believed to be stimulated by the presence of its substrate, lactose. In this view, populations that did not use milk as adults lost the ability to produce lactase, whereas groups that did consume milk and milk products retained lactase capability.

Biochemical studies cast doubt on this theory, and family studies have demonstrated that lactase production is controlled by an autosomal gene, recently located on chromosome 2. Persistence of lactase production is a dominant trait (Buller and Grand 1990; Arola and Tamm 1994). Following the terminology suggested by Gebhard Flatz (1987), the two alleles are designated LAC*P for lactase persistence and LAC*R for normal adult lactase restriction. The LAC locus appears to be a regulatory gene that reduces lactase synthesis by reducing the transcription of messenger RNA (Arola and Tamm 1994). Persons inheriting LAC*P from both parents would have lactase persistence into adulthood; those getting LAC*R alleles from both parents would display lactase restriction as adults. Heterozygotes would get different alleles and be LAC*P/LAC*R, but since LAC*P is dominant, lactase activity and ability to digest milk would persist beyond childhood.

Nutritional Implications

As milk and milk products are such rich sources of protein, calcium, carbohydrates, and other nutrients, the nutritional consequences of lactose intolerance in infants and children can be devastating, even lethal, unless other dietary sources are used. Formulas based on soybeans help many youngsters. Adults can get protein from other animal and vegetable sources or from fermented milk products. Yoghurt with live bacterial cultures may be tolerated well. Calcium can be obtained from dark green vegetables or from the bones of small sardines or anchovies consumed whole (Kretchmer 1993). It has been suggested that low milk consumption in elderly lactose intolerance adults might contribute to osteoporosis (Wheadon et al. 1991), but this has not been demonstrated. Lactose-free dairy products and oral lactase preparations are commercially available and help many people enjoy and gain the nutritional benefits of ice cream and other milk-based foods (Ramirez, Lee, and Graham 1994).

Lactase persistence is uncommon in Africans, Asians, southern Europeans, and the indigenous populations of the Americas and the Pacific. Questions have arisen concerning the use of milk as food for children. The American Academy of Pediatrics (AAP), noting the high nutritional value of milk for growing children, has determined that almost all U. S. children under 10, regardless of family background, can digest reasonable quantities of milk. The AAP recommends that the school-lunch half pint (about 240 milliliters [ml]) of milk be supplied to children up to this age, and notes that intolerance to 240 ml is rare even among older teens (American Academy of Pediatrics 1978). Similar results have been reported for African children in South African orphanages (Wittenberg and Moosa 1991). Malnourished African children, such as famine victims, also tolerate up to 350 ml of milk well, which allows the use of this valuable source of nutrients in emergency situations (O'Keefe, Young, and Rund 1990).

Testing for Lactase Persistence

Clinical diagnosis and population surveys for lactose digestion capabilities present several challenges. Clinical symptoms are discovered by self-reporting; thus, double-blind studies, in which neither the experimenter nor the subject knows if a challenge dose contains lactose or a placebo, are most useful. Direct lactase assay using biopsy specimens of intestinal mucosa is obviously an expensive and invasive method, practical only in particular clinical cases. Indirect assays require subjects to fast for several hours before being given doses of lactose in solution.

Then various tests are used to measure the splitting and subsequent metabolism of the disaccharide. Many of the older methods are cumbersome and imprecise. Blood samples may be tested for glucose before lactose challenge and at intervals afterward. High blood glucose levels after lactose ingestion indicate that lac-

tose is being split in the intestine. A variant of this method is to measure blood galactose. Since the liver metabolizes galactose, a dose of ethanol is given shortly before the experimental lactose to inhibit liver action. Another approach is to measure hydrogen gas excreted through the lungs. Subjects who cannot digest lactose will have hydrogen produced by colonic bacteria. Respiratory hydrogen can be conveniently and efficiently measured by gas chromatography. The ethanol-galactose and hydrogen methods are considered the most reliable; the hydrogen technique is cheaper, easier, and noninvasive (Flatz 1987).

Not only must the population studies of lactase activity be methodologically correct, the subjects must also be truly representative of their populations. Studies done on very small numbers of subjects or on hospital patients or other special groups may be unrepresentative. Indeed, some older studies may be unreliable due to poor techniques or sampling problems. Inter-marriage and genetic interchange also complicate analysis of the distribution of lactase persistence. Nonetheless, there has been great interest in the geographical and ethnic distribution of adult lactase persistence and the evolution of this unusual phenotype.

Distribution of Lactase Persistence

Several authors have compiled the results of regional studies (Flatz 1987; Kretchmer 1993; Sahi 1994). Some of the major findings are summarized here in Table IV.E.6.1.

It should be noted that data for northern India and Pakistan are suspect and that figures for Finno-Ugrian groups in northern Russia and western Siberia (Khanty, Mansi, Mari, Mordva, Nentsy) are based on small, possibly unrepresentative, samples and older methods (Valenkevich and Yakhontova 1991; Kozlov, Sheremeteva, and Kondik 1992). There is little hard information for the Balkan or Iberian peninsulas, Slavic territories east of Poland, Siberia, central Asia, or the Indian subcontinent. It would also be interesting to have more data on East African pastoralists, such as the Maasai, and on Baggara Arab and other cattle-keeping groups of the West African Sahel.

A high proportion of lactase persisters was noted in northwestern Europe in the early 1970s, and there were similar reports from northern India, from Bedouin and other pastoral populations in the Middle East and northern Africa, and from the Tutsi pastoralists of the Uganda-Rwanda region of East Africa. Very low rates were found among eastern, and most southern, Asians, most Africans, and native populations of the Americas and the Pacific, and only modest rates were found in southern and eastern Europe. In North and South America, Australia, and New Zealand, adult lactase ability is closely linked to place of origin; for example, white Australians resemble their European counterparts in lactase persistence, whereas Aborigines are almost entirely lactose intolerant. Varying degrees of Spanish and Indian

ancestry may explain regional differences in Mexico (Rosado et al. 1994). Similarly, a higher than expected prevalence of lactase persistence among Buryat Mongols of Russia's Lake Baikal region may be due to gene flow from European Russians (Kozlov et al. 1992).

Adult lactase capability appears to have evolved in two, and possibly three, geographic areas. The case is clearest and best documented for northern Europe, where there are very high percentages around the Baltic and North Seas. High levels of lactase persistence seem closely linked to Germanic and Finnic groups. Scandinavia, northern Germany, and Britain have high levels, as do the Finns and Estonians, the Finnic Izhorians west of St. Petersburg, the Mari of the middle Volga basin, and, to a lesser extent, their more distant relations, the Hungarians.

There is a general north-south gradient in Europe, which is evident within Germany, France, Italy, and perhaps Greece. As noted, more information is needed for Spain, Portugal, and eastern Europe, but there may be something of a west-east gradient in the Slavic lands. Varying frequencies of the LAC*P allele among Lapp groups may be related to differing lengths of historical use of reindeer and cow's milk and to admixture with other Scandinavians (Sahi 1994).

The second center of adult lactase persistence lies in the arid lands of Arabia, the Sahara, and eastern Sudan. There, lactase persistence characterizes only nomadic populations heavily dependent on camels and cattle, such as the Bedouin Arabs, the Tuareg of the Sahara, the Fulani of the West African Sahel, and the Beja and Kabbabish of Sudan. Lower rates among Nigerian Fulani may indicate a higher degree of genetic mixing with other peoples than among the Fulani of Senegal. In contrast, surrounding urban and agricultural populations, whether Arab, Turkish, Iranian, or African, have very low rates. It is interesting to note that the Somali sample also had a low frequency of the LAC*P allele. Possibly, pastoral Somali have higher prevalences than their urban compatriots.

A third center of adult lactase persistence has been suggested among the Tutsi population of the Uganda-Rwanda area of the East African interior. The Tutsi are an aristocratic cattle-herding caste of Nilotic descent who have traditionally ruled over agricultural Bantu-speakers. Table IV.E.6.1 shows that only 7 percent of a sample of 65 Tutsi adults were lactase deficient, but the data are old, there certainly has been some mixture with Bantu-speakers, and the study should be replicated. The Nilotic peoples of the southern Sudan, whence the Tutsi originated a few centuries ago, do not display this trait. Unless the Tutsi result can be confirmed, and the Maasai and other East African Nilotic groups can be tested, this third center of the LAC*P allele must be considered doubtful. If it does exist, it probably arose as a fairly recent mutation, as there are no obvious historical mechanisms to account for gene flow between the Tutsi and desert dwellers farther north.

Table IV.E.6.1. *Distribution of lactose phenotypes*

Country	Group	# High LDC	# Low LDC	% Low LDC	Frequency LAC*P allele ^a
Europe					
Sweden	Swedes	396	4	1	.90
Finland	Swedish-speakers	84	7	8	.72
Finland	Finnish-speakers	371	78	17	.58
Finland	Fisher Lapps	82	28	25	.50
Finland	Mountain Lapps	235	110	32	.43
Finland	Skolt Lapps	70	106	60	.23
Denmark	Danes	743	19	3	.83
Netherlands	Dutch	14	0	0	1.00
Britain	British	90	6	6	.76
Ireland	Irish	48	2	4	.80
France	Northern French	48	14	23	.52
France	Southern French	40	31	44	.34
Spain	Spaniards	226	39	15	.61
Germany	Germans	548	98	15	.61
Germany	Northwest Germans	405	36	8	.72
Germany	Northeast Germans	191	55	22	.53
Germany	Western Germans	157	25	14	.63
Germany	Bavarians	191	30	14	.63
Germany	Southwest Germans	104	32	23	.52
Austria	Austrians	422	106	20	.55
Czech Republic	Czechs	175	25	13	.64
Poland	Poles	172	103	37	.39
Hungary	Magyars	337	198	37	.39
Slovenia	Slovenes	99	54	35	.41
Italy	Northern Italians	191	192	50	.29
Italy	Southern Italians	31	78	72	.15
Italy	Sicilians	29	71	71	.16
Greece	Greeks	452	520	53	.27
Greece	Peloponesian Greeks	?	?	75	.13
Cyprus	Greeks	19	48	72	.15
Russia	Russians- St. Petersburg	210	38	15	.61
Russia	Russians	?	?	12.5–16.3	.60–.65
Russia	Izhorians	78	54	41	.36
Russia	Mordvinians	?	?	11.1	.67
Russia	Mari	?	?	81	.10
Russia	Mansi	3	9	75	.13
Russia	Khanty	3	7	70	.16
Russia	Nentsy	1	4	80	.11
Russia	Buryat Mongols	11	9	45	.17
Estonia	Estonians	467	183	28	.47
Latvia	Latvians	?	?	32	.43
Asia and Pacific					
Turkey	Turks	135	335	71	.16
Israel	Sephardic Jews	22	46	68	.18
Israel	Oriental Jews	29	65	69	.17
Israel	Ashkenazi Jews	11	42	79	.11
Jordan	Bedouin	123	39	24	.51
Jordan	Jordanians	43	161	79	.11
Syria	Syrians	7	68	91	.05
Saudi Arabia	Bedouin	12	2	13	.64
Egypt	Northern Egyptians	106	282	73	.15
Egypt	Central Egyptians	17	94	85	.08
Iran	Iranians	7	33	83	.09
Afghanistan	Pashtoons	15	56	79	.11
Afghanistan	Other Afghans	10	50	83	.08
Pakistan	Punjabi	132	190	59	.23
India	Northern Indians	194	70	27	.48
India	Central Indians	46	79	63	.21
India	Southern Indians	20	40	67	.18
Sri Lanka	Singhalese	55	145	73	.15
Thailand	Thais	0	149	100	0.00
China	Han	19	229	92	.04
China	Mongols	24	174	88	.06
China	Kazakhs	46	149	76	.13
China	Uighurs	26	16	38	.38
Taiwan	Chinese	0	71	100	0.00
Japan	Japanese	10	56	85	.08
Indonesia	Javanese	5	48	91	.05
Fiji	Fijians	0	12	100	0.00
Papua New Guinea	Various tribes	12	111	90	.05
Australia	Aborigines	48	97	67	.18
Africa					
Nigeria	Ibo, Yoruba	12	101	89	.06
Nigeria	Fulani	7	2	22	.53
Niger	Tuareg	103	15	13	.64
Senegal	Fulani	29	0	0	1.00
Somalia	Somali	58	186	76	.13
Kenya	Bantu-speakers	19	52	73	.15
Uganda, Rwanda	Bantu-speakers	14	100	88	.06
Uganda, Rwanda	Tutsi	65	5	7	.74
South Africa	Bantu-speakers	28	144	84	.08
South Africa	San (Bushmen)	3	62	95	.03
Sudan	Kabbabish nomads	39	12	24	.51
Sudan	Beja nomads	252	51	17	.59
Sudan	"Blacks"– farmers	20	64	76	.13
Sudan	Nilotics	72	210	75	.13
Americas					
Greenland	Eskimos	18	101	85	.08
Greenland	Mixed European-Eskimo	67	41	38	.38
United States	Native Americans	11	210	95	.03
United States	Blacks	138	252	65	.19
United States	Mexican- Americans	147	158	52	.28
Mexico	"Mexicans"	69	332	83	.09
Colombia	Mestizos	30	15	33	.43
Colombia	Chami Indians	0	24	100	0.00
Bolivia	Ayamara Indians	7	24	77	.12
Peru	Mestizos	26	68	72	.15

^aAssuming genetic equilibrium and calculated from the Hardy-Weinberg law.

Sources: Flatz (1987), Zheng et al. (1988), Valenkevich and Yakhontova (1991), Koslov, Sheremeteva, and Kondik (1992), and Kretchmer (1993).

Evolution of Lactase Persistence

Frederick J. Simoons (1969, 1970) has advanced the thesis that lactase persistence is closely linked to dairying. His culture-evolution hypothesis is that groups that kept cattle and other milk animals would gain a selective advantage if adults retained the ability to use milk and milk products as food. A mutation like LAC*P would be nutritionally beneficial, and the growing number of milk-using adults would then be encouraged to devote more effort toward livestock raising. In general, the distribution of adult lactase persistence and dairying shows a positive relationship. In areas with no dairying tradition, such as China, Oceania, Pre-Columbian America, or tropical Africa, few adults can digest lactose.

Northern Europe presents the opposite case. More data around the periphery of the two postulated centers would be highly desirable, and we know little about most of the stock-raising societies of central Asia. Still, although the correspondence is not perfect, and gene flow through population mixing complicates the picture, the association seems strong. Given the origins of cattle keeping about 4000 to 3500 B.C. in northern Europe, and even earlier in the Middle East, there probably has been enough time for modest selective pressures to have produced observed LAC*P rates (Sahi 1994).

Other selective forces may also have been at work. Flatz (1987) has suggested that calcium absorption was such a factor in northern Europe. Lactose is known to facilitate calcium absorption in the intestine. The cold, cloudy climate frequently discouraged skin exposure to sunlight, thereby reducing the body's production of vitamin D. Relatively little dietary vitamin D was available, and so in its absence, calcium was poorly absorbed. Northern populations were thus vulnerable to rickets and osteomalacia. Pelvic deformities made births more difficult. The gradual extinction of the Greenland Viking colony is an example; skeletal evidence shows that such bone diseases were common among this moribund population. A mutant LAC*P allele would not only allow adults to use an excellent source of calcium, but the lactose would also facilitate its absorption. While not proven, this hypothesis has attracted much attention. It would complement the theory that the pale skin of northern Europeans is a genetic trait maximizing the utility of sunlight in vitamin D production and, hence, calcium absorption.

Similarly, other selective pressures facilitating the survival of mutant LAC*P alleles have been postulated for the Sahara-Arabian Peninsula desert region. There is a high degree of dependency on milk among many groups of desert pastoralists, and so a positive link between lactase persistence and milking seems very plausible. In addition, it has been argued (Cook 1978) that the simple fact that milk is a liquid would give

adults who could consume it in large quantities a powerful selective advantage. The theory, while unproven, certainly seems plausible. G. C. Cook's suggestion that lactase persistence conveyed some resistance to gastrointestinal diseases has attracted much less support. At least for cholera, his claim must be rejected, based on what we know of the historical geography of the disease. Cholera seems to have been restricted to the Indian subcontinent until very recent times.

Finally, it seems most likely that the European and Arabia-Sahara centers of LAC*P prevalence, and the Uganda-Rwanda center (if it in fact exists), arose independently. Population movement and gene flow can be very extensive and, no doubt, have played a substantial role around the centers. Despite the efforts of some authors to find a common origin in the ancient Middle East, it is simpler to suggest independent origins than to postulate gene flow from the Middle East to Scandinavia and to the interior of East Africa. The problem might be resolved in the future if gene sequencing could show that the LAC*P alleles in Sweden and Saudi Arabia are, in fact, the same or are distinct forms of the gene with a similar function.

Conclusions

Lactose malabsorption is the normal condition of most adults. Many suffer the clinical symptoms of lactose intolerance if they consume milk, especially in large amounts. In two, or possibly three, places, genetic mutations have arisen that allow adults to gain the nutritional and culinary benefits of milk and many other dairy products. This ability has evolved along with cultural developments with profound implications for livelihood, including, in the northern European case, the development of mixed farming. East Asian, African, Oceanic, and Amerindian peoples, of course, thrived without this genetic trait and its cultural consequences. Their infants and young children enjoyed the nutritional advantages of milk; adults ate other things, including fermented milk products. Milk can be consumed by most lactose-intolerant older children in moderate amounts, and so milk can be a valuable nutrient for the undernourished or famine stricken. Modern commercial lactase products allow most lactose-intolerant adults to consume dairy products; thus, pizza and ice cream need not be forbidden foods.

Finally, the LAC*P and LAC*R genes are interesting far beyond their biomedical significance. Along with linguistics, archaeology, and physical anthropology, further research on lactase genes and other genetic markers will provide clues to the prehistory of peoples, their migrations and interminglings, and the origins and development of major language families.

K. David Patterson

Bibliography

- American Academy of Pediatrics, Committee on Nutrition. 1978. The practical significance of lactose intolerance in children. *Pediatrics* 62, no. 2: 240-5.
- Arola, H., and A. Tamm. 1994. Metabolism of lactose in the human body. *Scandinavian Journal of Gastroenterology* 202 (Supplement): 21-5.
- Buller, Hans A., and Richard J. Grand. 1990. Lactose intolerance. *Annual Review of Medicine* 41: 141-8.
- Castiglia, Patricia T. 1994. Lactose intolerance. *Journal of Pediatric Health Care* 8: 36-8.
- Cook, G. C. 1978. Did persistence of intestinal lactase into adult life originate on the Arabian Peninsula? *Man* 13: 418-27.
- Dahlqvist, A. 1977. The basic aspects of the chemical background of lactase deficiency. *Postgraduate Medical Journal* 53: 57-64.
- Flatz, Gebhard. 1987. Genetics of lactose digestion in humans. *Advances in Human Genetics* 16: 1-77.
- Kozlov, A. I., V. A. Sheremeteva, and V. M. Kondik. 1992. Issledovanie lactaznogo polimorfiza u predstavitelei razlichnykh etnoterritorialnykh grupp. *Biologicheskie Nauki* 1: 64-8.
- Kretchmer, Norman. 1993. Lactose intolerance and malabsorption. In *The Cambridge world history of human disease*, ed. Kenneth F. Kiple, 813-17. Cambridge and New York.
- O'Keefe, S. J., G. O. Young, and J. Rund. 1990. Milk tolerance and the malnourished African. *European Journal of Clinical Nutrition* 44: 499-504.
- Ramirez, F. C., K. Lee, and D. Y. Graham. 1994. All lactase preparations are not the same: Results of a prospective, randomized, placebo-controlled trial. *American Journal of Gastroenterology* 89: 566-70.
- Rosado, J. L., C. Gonzalez, M. E. Valencia, et al. 1994. Lactose maldigestion and milk intolerance: A study in rural and urban Mexico using physiological doses of milk. *Journal of Nutrition* 124: 1052-9.
- Sahi, T. 1994. Genetics and epidemiology of adult-type hypolactasia. *Scandinavian Journal of Gastroenterology* 202 (Supplement): 7-20.
- Simoons, F. J. 1969. Primary adult lactose intolerance and the milking habit: A problem in biological and cultural interrelations. I. Review of the medical research. *American Journal of Digestive Diseases* 14: 819-36.
1970. Primary adult lactose intolerance and the milking habit: A problem in biological and cultural interrelations. II. A culture historical hypothesis. *American Journal of Digestive Diseases* 15: 695-710.
- Tamm, A. 1994. Management of lactose intolerance. *Scandinavian Journal of Gastroenterology* 202 (Supplement): 55-63.
- Valenkevich, L. I., and O. I. Yakhontova. 1991. Rasprostranennost' defitsita laktazy sredi narodov SSSR. *Voprosy Pitaniya* 4: 10-13.
- Wheadon, M., A. Goulding, G. O. Barbezat, et al. 1991. Lactose malabsorption and calcium intake as risk factors for osteoporosis in elderly New Zealand women. *New Zealand Medical Journal* 104: 417-19.
- Wittenberg, D. F., and A. Moosa. 1991. The practical significance of lactose maldigestion in institutionalised black children. *South African Medical Journal* 79: 70-2.
- Zheng, Jia-ju, Gong Zeng-liang, Xue Lian-sheng, et al. 1988. Lactose malabsorption and its ethnic differences in Hans and Uyghurs. *Chinese Medical Journal* 101: 284-6.

IV.E.7 🐉 Obesity

Obesity is a dimension of body image based on a society's consideration of acceptable body size and, as such, is the focus of anthropological, sociological, and psychological study (de Garine and Pollock 1995). However, most of the research on obesity in Western societies has focused on medical issues ranging from genetic etiology to therapeutic interventions. Overfatness or obesity is a major health problem in countries that are affluent and is increasing in prevalence among the socioeconomic elite of those that are modernizing. An estimated 90 million Americans – one-third of the population – are substantially above their range of desirable body weight; in some other populations more than half of their members fit into this category.

Of course, some fat or adipose tissue is essential for life and serves a number of functions. It provides metabolic fuel; thermal insulation; a reservoir for vitamins, hormones, and other chemicals; and protection for the viscera and dermal constituents, such as blood vessels, nerves, and glands (Beller 1977). However, an excessive accumulation of fat is associated with an increased risk for diabetes, hypertension, cardiovascular and musculoskeletal problems, and in general, a reduced life expectancy. Moreover, in many societies, fatness elicits a psychosocial stigma.

Definitions and Diagnosis

Body weight is the most widely used anthropometric indicator of nutritional reserves, and weight relative to height is an acceptable measure of body size for growth monitoring and for most epidemiological surveys. Overweight and obesity, though often used synonymously, are not the same. S. Abraham and co-workers (1983) clearly made the distinction in analyzing data from the first U.S. National Health and Nutrition Examination (NHANES) survey. Overweight was defined as an excess in body weight relative to a range of weights for height. In this report, individuals over the 85th percentile of weight for height standards are considered overweight. Obesity was defined as an excess of body fat based on the sum of the triceps (upper arm) skinfold and subscapular (back) skinfold. Skinfold measurements using calipers that pinch a fold of skin and subcutaneous fat at specific sites (for example, waist, abdomen, thighs, upper arm, and back) are used in equations to estimate body fat stores and are compared with reference percentile tables (Himes 1991).

Many recent studies have used the Body Mass Index (BMI), which is the weight in kilograms divided by height in meters squared, to categorize body size. This index was devised by the Belgian mathematician Adolphe Quetelet (1796-1874) and is also referred to as the Quetelet index. (The Ponderal Index, which is the quotient of the height in inches divided by the cube root of the weight in pounds, has been similarly used.)

Overweight is defined as a BMI above 27.3 for women and 27.8 for men. These BMIs represent approximately 124 percent of desirable weight for men and 120 percent of desirable weight for women, defined as the mid-point of the range of weight for a medium-size skeletal frame from the 1983 Metropolitan Insurance Company Height and Weight Tables. The World Health Organization uses a similar range of BMIs: below 20 (lean), 20 to 25 (acceptable), 25 to 29.9 (moderately overweight), 30 to 39.9 (severely obese), and greater than 40 (morbidly obese). Epidemiological studies frequently use a BMI of 30 as the delimiter for obesity for both sexes.

Other anthropometric measurements have been used as alternatives to body weight in assessment of obesity. Body girth measurements or circumferences at specific anatomical locations have a high correlation with body mass. A commonly used measure is the circumference of the upper arm. This measurement, in conjunction with the triceps skinfold, has been used to compare the fat and lean components of the arm and thus to provide a measurement of energy and protein stores. More sophisticated, expensive, and time-consuming techniques assess the lean and fat components of the body. These techniques have included densitometry, magnetic resonance imaging (MRI), basic X rays, computerized tomography (CAT) scans, ultrasound, bioelectrical impedance, total body water, and body potassium levels (Lukaski 1987).

Skinfolds, circumferences, and imaging techniques assess the regional distribution of fat deposits. A central distribution of fat is referred to as an apple shape. A lower torso distribution of fat on the hips is referred to as a pear shape. The apple shape, often measured as a high waist-to-hip ratio of circumferences, is associated with internal deposits of abdominal fat and increased risk for coronary artery disease and adult onset diabetes. By contrast, the pear shape is not associated with increased disease risk (Bouchard and Johnston 1988).

Epidemiology

Current estimates of the prevalence of obesity indicate that it has reached epidemic proportions in some populations (Table IV.E.7.1). The most widely cited statistics on weight are those from NHANES III and are based on a random sample of the U.S. population between 1988 and 1991 in which 31 percent of males and 35 percent of females ages 20 to 74 years were considered overweight (Table IV.E.7.1). Table IV.E.7.2 presents the percentages of adults defined as obese with BMIs ≥ 30 . More alarming are the recent estimates of the percent overweight and obese done by the Institute of Medicine of the National Academy of Science (1995). Viewed in light of BMIs that are 25 or greater, 59 percent of American males and 49 percent of females are overweight or obese. Two percent of males and 4 percent of females are considered morbidly obese with BMIs over 40. A 5-foot 4-inch woman with a BMI of 40 weighs 230 pounds.

Table IV.E.7.1 *Prevalence of overweight (1980s-90s), based on Body Mass Index (kg/m²) or weight for height references*

Population	Ages	Males (%) ^a	Females (%) ^a
North America			
United States population	20-74	31	35
Non-Hispanic white	20-74	32	34
African-American	20-74	31	49
Mexican-American	20-74	36	47
Seminole Native American	Adult	>50	>50
Pima	Adult	>50	>50
Zuni	Adult	29-35	55-66
Canadian population	25-64	9	8
Canadian Cree & Ojibwa	Adult	45-54	>90
South Pacific			
Micronesian	Adult	85	93
Polynesian	Adult	48	79
Melanesian	Adult	31	65
Asian	Adult	12	36
Native Hawaiian	Adult	85	52
Europe			
English	18-84	13	15
Italian	45-64	10	11
Finnish	50-9	12	30
Dutch	50-64	5	10
Australia	25-64	7	9
South Africa	15-84	15	18
South America			
Costa Rican	40-5	8	14
Salvadoran	40-5	0	2
Guatemalan	40-5	0	6
Nicaraguan	40-5	3	16
Panamanian	40-5	2	2
Honduran	40-5	3	6

^aRounded to the nearest percentage point.

Sources: Data are from many sources.

A survey of Micronesian Islanders indicates that 85 percent of males and 93 percent of females are overweight, whereas among native Hawaiians, a Polynesian group, 85 percent of males and 62 percent of females are overweight. Obesity is also prevalent in a number of native North American groups. A survey of Seminoles and Pimas has revealed that more than 50 percent of the adults are obese, whereas among the Canadian Cree and Ojibwa more than 90 percent of females and between 45 and 54 percent of males were so categorized. Among other ethnic groups within the United States there also are high levels of adult obesity: In Texas, fully 66 percent of male Mexican-Americans and 60 percent of females are obese, as were almost 50 percent of female African-Americans and 31 percent of males nationwide. Table IV.E.7.1 and Table IV.E.7.2 present the proportion of obese or overweight adults in a number of countries.

Table IV.E.7.2. *Prevalence of obesity**
(Body Mass Index ≥ 30 kg/m²)

Country/region	Age	Percentage males	Percentage females
United States			
White	40-9	14.8	16.4
Black	40-9	23.1	33.0
Hispanic	40-9	18.5	38.7
Pima Indian	35-44	64.0	75.0
Canada	35-44	12.0	16.0
France	16-84	7.0	8.0
Sweden	15-44	4.8	3.9
Italy	45-64	9.9	11.1
Netherlands	35-49	4.2	5.0
United Kingdom	35-49	7.9	8.6
Germany	25-69	16.0	16.0
Costa Rica	40-5	5.7	14.4
El Salvador	40-5	0.0	1.5
Guatemala	40-5	0.0	5.6
Honduras	40-5	2.8	6.0
Australia	25-64	8.6	13.3
South Africa			
Cape Peninsula	35-44	14.3	15.6
Solomon Islands			
Urban	35-54	19.0	43.0
Samoans (Hawaii)	45-54	45.0	55.0
India (Bombay)	15-76	6.2	10.7
Thailand	over 30	11.1	11.1

*Data are from 1980s.

Source: Bouchard (1994), pp. 27-30.

Obesity can begin in early life. Eleven percent of U.S. children (ages 6 to 11 years), 13 percent of adolescent males (ages 10 to 17), and 9 percent of adolescent females are overweight. Outside of the United States, M. Gurney and J. Gorstein (1988), who surveyed the preschool populations of 34 countries, found ranges of obesity from 1 to 11 percent. By way of a few examples, in Jordan and Tahiti 2 percent of the preschoolers were obese, in the United Kingdom 3 percent, in Canada 6 percent, and in Jamaica the figure was 10 percent. There are no comparable multinational studies for school-age children.

Obesity in childhood and adolescence is a good predictor of obesity in adulthood. In one study, a third of the obese adults examined were already overweight or obese at 7 years of age, and two-thirds were overweight or obese by age 14. Studies in the United States and Britain found that between 40 and 74 percent of obese 11- to 14-year-old youngsters became obese young adults.

Secular Trends in Overweight

Since 1960, surveys in the United States have tracked changes in the proportion of those overweight and obese in the population. In general, adult average weight and the proportion of individuals who are overweight have grown larger. Overweight preva-

lence for adults increased 8 percent between the recording periods of NHANES II (1976-80) and NHANES III (1988-91), as shown in Table IV.E.7.3. The mean BMI jumped from 25.3 to 26.3 and the mean body weight increased by 3.6 kg. The proportion of overweight children and adolescents, in particular, has been augmented, although a countertrend can be seen among the older (over 70 years) segment of the population. Thus for women 60 to 74 years of age, there was a decrease in the percent overweight from 45.6 percent to 41.3 percent. However, among African-American females, corresponding figures for the younger cohort (50 to 59 years) showed an increase from 35 to 52 percent overweight. By contrast, African-American males age 60 to more than 80 years had the lowest percent of overweight individuals in comparison with white and Hispanic Americans. It is notable that during the two decades from 1960 to 1980, no consistent secular trends were found for whites or blacks ages 12 to 17 years and 18 to 34 years. However, this changed significantly with the NHANES III data, which revealed increasing obesity in both of these groups.

Based on self-reports in a Harris Poll, 74 percent of Americans ages 28 and older stated in 1996 that they were overweight. This was an increase from 71 percent in 1995, 69 percent in 1994, and 59 percent in 1986. In the United Kingdom during the decade 1980 to 1990, the percentage of adult males classified as overweight grew from 39 to 48 percent and those who were obese from 6 percent to 8 percent. The proportion of women classified as overweight increased from 32 percent to 40 percent and obese from 8 percent to 13 percent. Overweight percentages were highest for males and females 50 to 64 years of age.

Gender

Childbearing and menopause are associated with weight gain, obesity, and an increasing waist-to-hip ratio. Maternal body fat is gained during pregnancy in response to the hormonal milieu, with a third of women gaining more than 5 kilograms of adipose tissue. In the United States, the mean net weight gain

Table IV.E.7.3. *Age-adjusted and age-specific prevalence of overweight (1960-91)*

Population groups	NHES 1 (1960-2)	NHANES I (1971-4)	NHANES II (1976-80)	NHANES III (1988-91)
White males	23.0	23.8	24.2	32.0
White females	23.6	24.0	24.4	33.5
Black males	22.1	23.9	26.2	31.8
Black females	41.6	43.1	44.5	49.2

Note: NHES = National Health Examination Survey

NHANES = National Health and Nutrition Examination Survey

Source: Kuczmarski et al. (1994), pp. 205-11.

with each childbearing cycle is 1 kilogram above that normally gained with aging. Americans put on approximately 20 pounds from age 25 to 55. Lactation mobilizes fat, but selectively from the femoral region, and, therefore, there is still an increase in the waist-to-hip ratio. Menopause has also been reported to increase the waist-to-hip ratio and add an average 20 percent body fat mass, compared to the premenopausal state. Both subcutaneous and internal visceral abdominal fat increase in postmenopausal women.

Obesity is approximately twice as prevalent among women as men in the United States, although this disparity is the most striking in certain populations, such as African-Americans, Mexican-Americans, Puerto Ricans, and Western Samoans (Tables IV.E.7.1 and IV.E.7.2). By contrast, Hawaiians, Nauruans, Native Americans, Alaskan natives, and Mexican-Americans have the highest obesity prevalence among males. Obesity is only slightly more prevalent among black, Puerto Rican, and Cuban-American men than among non-Hispanic white men. Asian-Americans (of Chinese, Japanese, Filipino, and Indo-Chinese origin) have a lower obesity prevalence than other minority groups in the United States, although this may be changing. Some groups, such as the California Japanese, have recently developed moderately high BMIs.

P. Brown and M. Konner (1987) have noted that females appear to become obese with modernization. They suggested that the sex ratio of obesity is a marker for a population on a trajectory of economic development and westernization. That is, an excess of female versus male obesity is more likely to be observed in poorer populations in the developing world and less so in affluent Western populations. The data are generally consistent with this interpretation.

Brown's (1991) cross-cultural survey using the Human Relation Area File data found that 81 percent of societies for which there was sufficient data rated "plumpness" or being "filled out" as an attribute of beauty in females. This was particularly true of fat deposits on the hips and legs. Bigness for women in some groups is a sign of power, beauty, and maternity. Indeed, anthropologists have described the practices of populations in Polynesia (Pollock 1995) and West Africa (Brink 1995), where young women are secluded for a year or more in "fattening huts" prior to marriage. Such plumpness is not only considered desirable in a woman but also reflects positively on the socioeconomic status of her family and its ability to feed a daughter without having to rely on her labor.

Modernization

The construct of modernization encompasses a wide variety of lifestyle changes, including physical activity patterns, diet, and psychosocial stress. Many "diseases of civilization" (Trowell and Burkitt 1981) have been associated with westernization or modernization.

For example, the Pimas of Arizona have the highest rate of diabetes of any known population – a condition that is accompanied by a high prevalence of obesity among individuals of all ages. However, this is not the case with a small group of Pima whose ancestors migrated to Mexico some 700 to 1,000 years ago and who live today in a remote, mountainous location with a traditional lifestyle that is in marked contrast to the Arizona Pima. These Mexican Pima are lighter in weight and shorter and have lower BMIs, plasma cholesterol levels, and rates of diabetes. Consequently, it would seem that much of the problem with the Arizona Pima lies in lifestyle. Suggestive as well was a study conducted by K. O'Dea (1984), who took a group of diabetic Australian Aborigines away from an urban lifestyle to live as hunters and gatherers for seven weeks in northwestern Australia. The subjects lost an average of 8 kilograms and experienced improved carbohydrate metabolism.

In another study, the situation was reversed, whereby 13 Tarahumaras living a traditional lifestyle in northern Mexico were fed a diet typical of affluent societies for five weeks. On average, these subjects gained 3.8 kilograms (or 7 percent of their initial body weight) and had dramatic increases in plasma lipids and lipoprotein levels. Clearly, such investigations indicate that aspects of modern lifestyle can significantly contribute to obesity as well as other deleterious metabolic changes. They point to the benefits of a diet low in animal fat and high in complex carbohydrates, as well as the importance of high levels of physical activity (McMurphy et al. 1991).

In addition, migration studies have shown that populations moving from traditional to westernized environments experience large increases in body weight, along with rising rates of diabetes and other metabolic changes (Bindon and Baker 1985; Bindon 1995).

Geography

Derek Roberts's study (1953) of the geography and climate of 220 societies revealed that height and weight ratios are related to mean annual temperatures and that people are fatter the farther away they live from the equator. In other words, populations are fattest where summers are the coldest and leanest where summers are the hottest (Beller 1977). Put another way, if height is held constant, heavier people are found in the world's colder climates. Roberts hypothesized that cold stimulates the adrenal glands which, in turn, increase fat deposits.

In the United States, obesity is most common in the Northeast and Midwest, and rates are significantly higher in metropolitan regions than in rural areas. However, cutting across geographic distinctions are ethnic group concentrations and socioeconomic classes. In modernizing countries, the more affluent segments of populations exhibit a higher prevalence of obesity regardless of rural or urban location. But in

industrialized countries, the lower socioeconomic classes have the higher prevalence, and in addition, immigrant populations show increases in body weight compared to their sedentary counterparts or populations of origin. Finally, within a country, migration from rural to urban areas also leads to obesity.

Lifestyle Factors

Physical Activity

Physical inactivity has been related to increases in body weight and to obesity in both children and adults. Research by William Dietz and S. L. Gortmaker (1985) has demonstrated a linear relationship between the number of hours of television watching and body weight among Americans. Moreover, cross-sectional studies indicate that there is a negative relationship between energy expended for physical activity and body fat content. However, these associations do not prove that low levels of physical activity promote high levels of body fat because the association can also mean that individuals with existing high levels of body fat are rendered unable to exercise vigorously or for extended durations.

Both aerobic exercise and resistance training result in decreases in body fat by increasing energy expenditure during the actual period of exercise and subsequent periods of rest. Thus, exercise can promote a negative energy balance provided that there is not compensatory energy intake for those calories expended during and after exercise. Exercise has been found to be moderately successful in promoting and maintaining weight loss, and in some studies, weight loss was sustained with moderate exercise after the cessation of dieting.

Diet

M. Suzuki and N. Hosoya (1988) have succinctly modeled the relationship between obesity in adulthood and dietary changes, by arguing that modernized diets and food habits tend to accelerate the storage of body fat, regardless of the amount of energy that is ingested. Involved are (1) decreased carbohydrate and increased fat proportions in daily energy intake; (2) gorging just before resting; (3) increase of simultaneous intake of fats and sugars; (4) increased consumption of cereals in refined forms, such as flour, rather than unrefined, unprocessed grains; (5) increased consumption of soft, digestible foods rather than hard and more difficult to digest (for example, fibrous) foods; and (6) increased consumption of alcohol.

Although the primary focus has been on caloric intake, data also indicate that the composition of the diet is important in terms of metabolic rate, energy storage, and the production of obesity. Both the amount and the composition of food influence body-weight regulation. In a clinical study by T. Horton and colleagues (1995), carbohydrate overfeeding produced progressive increases in oxidation and total

energy expenditure, resulting in 75 to 80 percent of excess energy being stored. Fat overfeeding had a minimal effect on oxidation and total energy expenditure, leading to storage of 90 to 95 percent of excess energy. Excess dietary fat led to a greater fat accumulation than did excess dietary carbohydrates. Other investigations have demonstrated that individuals with a family history of obesity are more likely to suffer the obesity-promoting effects of high-fat diets than individuals without such a family history (Bouchard and Bray 1996).

Epidemiologic data from both the United States and Great Britain have shown that sugar intake is inversely related to obesity prevalence but that BMI and the percent of calories from fat are positively correlated. Indeed, a number of studies indicate that it is dietary fat, rather than sugar intake, that promotes obesity. However, a high-carbohydrate intake may cause hyperinsulinemia that, in turn, promotes fat storage. Excess intake of any macronutrient can contribute to weight gain, especially when associated with low energy expenditure.

The nutritional epidemiological transitions that promote obesity are, in part, fueled by multinational food corporations, which have introduced calorically dense foods that are advertised widely. The U.S. Department of Agriculture estimates that in the United States alone, the food and restaurant industry spends approximately 36 billion dollars annually on advertising, and some campaigns for single items (for example, a new soft drink or a new hamburger) exceeded 50 million dollars in 1996 and 1997. The trend in the United States is to increase the size of fast-food items (for example, “supersize,” or “giant size”), which are generally high in fat, carbohydrates, sodium, and calories to begin with. The larger portions tend to be eaten as quickly and completely as the regular ones. Thus, McDonald’s “supersize” serving of fries, which contains 540 calories, represents a 20 percent increase in calories over their “regular” serving. A “king size” candy bar may be as much as 80 percent larger than the regular size bar, and a large popcorn in a movie theater is 50 percent bigger than a medium size. In the United States, a mean of 3,700 kilocalories are available each day for every man, woman, and child, representing a third more than the recommended dietary allowance for men and twice that for adult women.

The explosion of “light” and “low fat” and “fat free” foods has led consumers to believe that these items are also calorie reduced or even calorie free, although their caloric content may be equal (or nearly so) to the nonspecialized product. (One venerable exception is the diet soft drink, and the new calorie-free fat substitutes now entering the market may constitute others.) Salad bars, also popular in the United States, can be equally deceptive. Consumers put high-calorie salad dressings on low-calorie salads to the extent that the caloric content of the salad exceeds that of a

meal containing animal protein and fat. American women 19 to 50 years of age get more fat from salad dressings than from any other food (Hurley and Collins 1997).

Obesity is promoted not only by the consumption of fat but also by the overconsumption of carbohydrates and protein. Americans, in particular, are consuming too many of these macronutrients while maintaining or insufficiently reducing their intake of fats. Data from NHANES I (1970s) and NHANES II (1976–80) indicated that Americans only reduced their fat intake from 42 percent to 38 percent of calories – still far above the recommended fat intake of 30 percent of total daily calories (Bray 1993b).

Smoking

Smokers weigh consistently less than nonsmokers, and when individuals stop smoking they generally gain weight. Ex-smokers reach body weights similar to those of age- and sex-matched nonsmokers, although gross obesity appears to be more frequent in ex-smokers than in those who have never smoked. The increases in body weight of ex-smokers stem from a number of causes, including increased food consumption and decreased metabolic rate. In one study, it was found that young adults who smoked 24 cigarettes per day had a 200 kilocalorie greater daily expenditure of energy than when not smoking, and this increased energy expenditure was independent of energy intake and physical activity (Hofstetter et al. 1986). Another study revealed that middle-age and older male smokers had higher waist-to-hip ratios than nonsmokers, after controlling for BMI, dietary and alcohol intake, and activity levels (Troisi et al. 1991).

Socioeconomic Factors

Many studies have demonstrated a striking inverse relationship between socioeconomic status and the prevalence of obesity, particularly among women in developed countries. This relationship is true regardless of whether socioeconomic status (SES) is based on family income, educational level, or occupation. Fully 30 percent of women of lower SES in the United States are obese compared with less than 5 percent of those of the upper status groups. Some investigations have demonstrated that upwardly mobile women are less obese than women who remain in a low SES. The prevalence of obesity for men in lower SES is 32 percent compared with 16 percent among upper-class men.

Obesity is a socioeconomic disability in Western cultures. In general, the theory is that socioeconomic status influences obesity by education, income, and occupation, causing variations in behavior that change energy consumption and expenditure. However, obesity influences socioeconomic status by the stigmatization and discrimination it elicits, which, in turn, limits access to higher SES roles. There are

ample data showing discrimination against obese individuals in terms of access to education, hiring for a variety of occupations, salary, and advancement (Allon 1982; Cassell 1995).

In a pioneering study, J. Sobal and A. J. Stunkard (1989) looked at the relationship between socioeconomic status and obesity in both developed and developing societies. They found that there was an inverse relationship between SES and obesity in industrialized societies, even stronger for women than for men, but a direct relationship in developing societies. The data pertaining to children are less clear. For boys in industrialized societies, the relationship between SES and obesity was either inverse or absent, but in developing countries it was clearly direct. For girls in industrialized countries, the relationship was inverse, as it was for women, whereas no relationship was found in developing countries.

Income is related to obesity mostly through access to resources. Individuals and families with higher incomes have more options in terms of access to food and food choices, although actual caloric intake may not vary by income. Occupation is related to obesity primarily through lifestyle factors in terms of energy expenditure on the job and during leisure activities. And finally, educational levels have been related to the prevalence of obesity, as lower educational levels are associated with lower income.

History of Obesity

Anthropological constructions have indicated a hunter-gatherer lifestyle during most of human history that was marked by much physical activity to secure adequate food, interspersed seasonally with decreased food intake. This variation selected genetically for individuals who were able to store energy as fat to carry them through lean times. Both contemporary foraging populations and those engaged in the incipient domestication of plants and animals show seasonal changes in weight that reflect variations in the availability of foods.

Studies of traditional hunting and gathering populations report no obesity. In contrast, many examinations of traditional societies undergoing processes of modernization that include production of generally high-carbohydrate food crops demonstrate a rapid increase in the prevalence of obesity. In fact, H. C. Trowell and D. P. Burkitt (1981) have noted that obesity in modernizing societies is the first “disease of civilization” to appear. The rapidity with which obesity becomes a health problem in the modernization process highlights the critical role of behavioral factors in its causation, as well as the evolutionary genetic propensities that were adaptive for traditional, calorically expensive lifeways.

Given the rarity of obesity in preindustrial societies, it is not surprising that there is a lack of ethnomedical terms for the obese state (Brown 1991). In

fact, thinness has often been seen as a symptom of starvation or as a sign of disease, whereas plumpness has been viewed as a marker of health. For example, the Tiv of Nigeria distinguish between the very positive category of “too big” and the unpleasant condition of “to grow too fat.” In some societies, for women in particular, fatness has been and remains a symbol of maternity and nurturing. The concepts that fat babies and children are healthy and that food is a symbol of love and nurture are nearly universal (Brown and Konner 1987). Moreover, from an evolutionary perspective, there are many biological advantages to the maintenance of energy stores as fat. These include an ability to survive longer during a fast, a greater ability to fight infectious diseases, fewer gastrointestinal tract problems, less anemia, healthier and higher birth-weight babies, and earlier age of menarche (Cassidy 1991).

The ethnographic data concerning body preferences in males is relatively weak, although it does suggest a preference for a more muscular physique, moderately tall stature, and general largeness. Traditionally, big, but not necessarily obese, men have been seen as successful. However, historical trends show variations in positive and negative associations with obesity.

Historical Medical Concepts

George Bray (1992) has outlined the scientific and medical history of ideas concerning obesity. In the Hippocratic texts, obesity was associated with infertility in women, a laxity of muscle, a red complexion, and sudden death. To lose weight meant strenuous physical activity before consuming meals prepared with sesame seasoning and fat because these satiated appetite. Dieters were to eat just once a day, take no baths, sleep on a hard bed, and walk naked as long as possible. Galen (130–215), who followed in the Hippocratic tradition, identified types of obesity and prescribed bulky foods with low nutrient content, baths before eating, and vigorous exercise.

The first monographs in which obesity was the primary subject were published in the seventeenth century. A mechanistic model of the body was popularly invoked, although there were other theories about fatness based on fermentation and putrefaction as the basis for an iatrochemical model.

The medical history of the eighteenth century was dominated by Hermann Boerhaave (1688–1738), who has been called the “most successful clinician and medical teacher of the century” (Ackerknecht 1982: 130), although his influence was exercised mostly through the students he trained. During the first part of the century, the earliest English language monograph on obesity and 34 doctoral dissertations dealing with the subject were published. A common theme was the imbalance of various systems resulting in disease. For example, the essay by Thomas Short (1690–1772) – a pioneer in vital statistics – on the origin of corpulence attributed it to blood stored in the

oily parts of the body and not sufficiently discharged by perspiration. Short thought that fat was stored in little bags. He noted that corpulence was more common in countries with “wet air,” which he believed decreased perspiration. Foods that were soft, smooth, sweet, and oily, as well as slothfulness, led to obesity. Thus, exercise was viewed as important, as were diets light in foods of a “detergent kind.” Less nutritious kinds of food, such as fish, were to be consumed sparingly. Also recommended was less sleep, a reduction in passions, gentle evacuations, and tobacco smoking to stimulate the nerves of the mouth (Bray 1992).

During the second half of the eighteenth century, Boerhaave’s students dispersed from Leyden to found new centers of clinical medicine at Edinburgh and Vienna, and during these years obesity, corpulence, and polysarcia became “species” following a Linnaean system of classification.

In the aftermath of the French Revolution, a new vitality emerged in clinical medicine, stimulated by the Paris Clinical School, as well as in basic sciences. The concept of energy balance developed, following the work of Antoine L. Lavoisier (1743–94) and the elaboration of the law of thermodynamics by Hermann L. F. von Helmholtz (1821–94). T. K. Chambers wrote extensively on obesity using theoretical models derived from thermodynamics.

English clinical medicine in the nineteenth century is notable for a work by W. Ward entitled *Comments on Corpulency, Liniments and Leanness* (1829), which describes a number of clinical cases of massively obese individuals. In 1863 William Banting (1779–1878) penned the first popular diet pamphlet. Banting, a London undertaker, who had personally lost considerable weight to regain his health, advocated a diet of lean meat, dry toast, soft-boiled eggs, green vegetables, and liquids (Bray 1992).

The adipocyte (fat cell) was identified in German laboratories, and advances in neurology led to the description of several hypothalamic-pituitary causes of obesity, such as the Fröhlich syndrome (1901) and the Prader-Willi hyperphagia syndrome (1956); the Pickwickian hypoventilation syndrome (1956) is a nonneurologic type of obesity-associated ailment (Burwell et al. 1956; Butler 1990).

In the twentieth century we have seen an ever-growing understanding of the psychological, social, and neurophysiological mechanisms that control food intake. We have also just begun to develop important pharmaceutical interventions targeted at pathologies within these mechanisms.

Obesity in America

Hillel Schwartz (1986) and Jo Anne Cassell (1995) have both reviewed the history of obesity in the United States. Although many Americans currently associate body weight and size with moral weakness, such negative perspectives do not have a long history, and in fact, attitudes toward overweight people have

changed many times throughout history. In general, when food was scarce, such individuals were viewed as prosperous and envied by their neighbors. By contrast, in times and societies with ample food, fashion usually favored slim and lean figures. In many instances, however, no moral judgments were attached to either overweight or lean status.

Beginning in the medieval period and continuing in America throughout the eighteenth century, morality plays were popular entertainment. Gluttony, one of the cardinal sins, frequently appeared as a theme involving excess or greed – and certainly a lack of self-restraint with regard to food and drink. The sin, however, was in consuming more than one needed and leaving others with less than their share. Body size or shape were not factors in gluttony, and being overweight carried no implication of sin.

In the early part of the nineteenth century, medicine focused on appetite and not thinness, and most physicians believed that fat represented a reserve that could be called upon in the event of disease, trauma, or emergency. Beginning at about mid-century, however, attitudes began to change under the influence of such individuals as Sylvester Graham (1794–1851) and the Kellogg brothers, who advocated a vegetarian regimen, coarsely milled flour, the consumption of cereal grains, and moderation in diet. They declared gluttony an evil, and obesity became a moral issue. During the Victorian period, however, in a counter-movement, fashion again began favoring a rounded body shape and seven-course family dinners. Diamond Jim Brady and his companion, Lillian Russell, were widely admired for their insatiable appetites and came to symbolize the exuberant excesses of the era (Cassell 1995).

By the turn of the twentieth century, scientists had come to believe that body fat had its origin in the fat of foods consumed and that dietary fat passed unchanged through the digestive tract to be absorbed and deposited. Fat was again out of fashion. Cartoons and jokes about overweight persons began to appear in newspapers and magazines. Stout, once a perfectly good word, became uncomplimentary, and a person who was fat was considered ugly. William Howard Taft, at 6 feet 2 inches tall, weighed 355 pounds when he was President of the United States (1909–13). Newspaper cartoons showed his rotund size, and when he got stuck in the White House bathtub, the event was well publicized. A variety of products to help in weight loss suddenly became available, including appetite suppressants, diuretics, stimulants, and purgatives. Special teas, bath salts, and mechanical devices were also employed (Schwartz 1986).

A 1912 study by actuaries of insurance policy holders provided height, weight, and mortality figures that became the data base for actuarial tables used to determine, for the first time, that there was a relationship between body weight, health, and mortality. According to these statistics, moderate weight gain was appropri-

ate before age 35 but became increasingly harmful in later life. Excess body fat had become a serious health liability, and during World War I, the entire nation went on a diet. Rationing and conservation efforts designed to ensure adequate rations for the soldiers also conveyed the notion that it was patriotic to be thin; to carry excess weight was un-American.

Obesity researchers in the 1920s and 1930s were divided as to whether obesity was due to exogenous or endogenous causes. “Exogenous” people overate but had firm muscles and were basically in good health. They were cheerful and happy. “Endogenous” people had a deficient metabolism, flaccid muscles, and poor health and were sad or sour in nature. Not surprisingly, men were generally placed in the former group and women in the latter. For some researchers, weight was a genetic matter, with fatness a dominant trait and slenderness a recessive trait.

As early as 1911, thyroid supplements had become available and were regularly prescribed for “glandular” disorders. Indeed, for a brief period in the postwar years, American physicians thought that many obese patients were suffering from inadequate amounts of thyroid hormones. So long as it seemed that there were endocrinologic and genetic causes for obesity, there was once again an abatement of moral judgment. But in the 1930s, medicine concluded that few people had true thyroid deficiencies after all, and nutritional scientists had reached the position that people were overweight simply because they ate too much. Moral judgments returned, and obesity was recast as the outcome of psychological problems leading to overeating.

The introduction of motion pictures with their frequently svelte movie stars led, among other things, to the popularity of weight-loss diets, particularly those originating in California. Perhaps the most famous was the “Hollywood 18-Day Diet.” Another was the “banana diet” – the product of bananas and skim milk in a blender – and a forerunner of today’s liquid diets.

Research emphasis changed once again in the 1940s as scientists defined obesity as “over-fatness.” People learned that they might be overweight, but not overfat, and dieticians began to measure body fat with skinfold calipers. Psychological theories focused on such issues as depression, lack of self-esteem, boredom, and inner emptiness as precursors to overeating, which also acted as popular explanations. Fatness as a signal of psychological distress continues to be influential in many weight-loss programs that employ support groups.

In the 1950s, changes in American society brought more attitudinal changes about dieting and weight control. Supermarkets began to offer a number of low-calorie and diet products. Among these were 900-kilocalorie-per-day liquid-diet formulas that were an instant success and continue to be widely available in the United States. Surveys reported that 40 percent of all families were regularly using low-calorie or diet

products by 1962 and 70 percent by 1970 (Cassell 1995).

In today's marketplace, novel fat substitutes are occupying shelves along with earlier sugar substitutes. New products from cookies to ice cream are both fat and sugar free. Books on dieting have also been popular, and in the 1960s, those on weight loss were best-sellers, especially if the authors promised that loss without the need to give up one's favorite foods. Five million copies of *The Doctor's Quick Weight Loss Diet* (Stillman and Baker 1967) were sold in 1967 alone, and books on dieting continued to lead book sales in the 1990s. In addition, in 1972 psychological therapy started to employ behavioral therapy, which has continued as a popular approach both for diet groups and commercial weight-loss centers. How, why, and where persons ate became as important as what they ate.

In the 1980s, Americans began to focus on physical fitness and exercise as a way to promote good health. A plethora of articles on dieting appeared in newspapers and popular magazines, along with information on increased activity levels. The message was to lose weight with exercise and diet, so as to be happier, healthier, and more attractive. A diet industry mushroomed to assist Americans in losing weight. The number trying to do so was estimated to be one-fourth of men and one-half of women, and annual sales of diet products had exceeded 30 billion dollars by the end of the 1980s.

Nonetheless, the incidence of obesity in America continued to increase and, predictably, there has been something of a backlash. In today's environment some overweight individuals have used the courts to press cases of discrimination based on body size, and there is a growing antidiet movement, with fashion houses, magazines, and psychological and educational support groups that champion the obese. These groups, such as the National Association to Advance Fat Acceptance (California), Ample Opportunity (Oregon), and the Diet/Weight Liberation Project (New York), promote positive images of obese individuals, arrange for social gatherings with singles groups, and work toward equity at places of employment and other venues where there is discrimination against obese individuals.

Many groups have newsletters; there are a number of national magazines, such as *Radiance: The Magazine for Large Women* (Oakland, Calif.), and some presses specialize in publications about obesity, food, eating disorders, and related psychosocial, political, and medical issues, as, for example, *Fat Liberation Publications* (Cambridge, Mass.). Other inroads are being made in the media and on educational and economic fronts to reduce the stigma of obesity in the United States. Given the high percentage of overweight Americans, the tide may once again be turning.

Nonetheless, Cassell (1995) suggests that weight watching and concern for body image have become

an integral part of American culture as we enter the twenty-first century, and there is still great pressure to be thin. The health risks of excessive body fat are well documented scientifically, and articles on the subject appear daily in the popular literature. At the same time, a person's overweight status does not automatically imply ill health, and more effort on the part of health professionals is being directed toward evaluating an individual's overall health status before a recommendation for weight loss is made.

Causes of Obesity

Obesity has a multifactorial etiology, which includes genetic factors, metabolic and behavioral phenotypes, and environmental agents. Because it is a condition of excessive storage of energy, an understanding of energy balance is fundamental to understanding obesity. Energy balance is equal to energy intake minus energy expenditure, and, consequently, obesity becomes the result of a positive energy balance. Yet, energy balance is exquisitely sensitive. The average nonobese American male consumes approximately 1 million kilocalories per year, but his body fat stores remain unchanged if he expends an equal number of calories. However, a change of only 10 percent either in intake or output can lead to a 30-pound weight change in a single year (Bray 1987). More subtly, a gain of 11 kilograms (24 pounds) of weight during a 40-year time span can come about with a mean daily discrepancy between intake and expenditure of only 5 kilocalories.

Energy Intake

Individual differences in metabolic mechanisms are not well understood. Cross-sectional studies reporting energy intake for individuals with different body compositions have found that obese individuals may have high, normal, or even low energy intakes relative to normal-weight subjects (Lachance 1994) and that there is a poor correlation between daily energy intakes and expenditures (Edholm 1973). For most adults, the sensitivity of the energy balance system for change is less than 1 percent per year. The "normal" adult contains 140,000 kilocalories of energy in body fat, 24,000 kilocalories in protein, and only about 800 kilocalories in carbohydrate. Consequently, although an individual consuming 2,000 kilocalories per day of which 40 percent is carbohydrate will ingest an amount of carbohydrate comparable to body stores, protein intake will average only about 1 percent of total stores and fat intake considerably less than 1 percent (Bray 1987, 1993a, 1993b).

Energy balance with regard to the macronutrients has been illuminated in recent years by the discovery of specific enzymes and neurotransmitters with receptors in the central nervous system. The brain is sensitive to changes in circulating glucose levels, and a glucostatic mechanism may regulate the intake of

fat and carbohydrates – the primary energy substrates. Carbohydrate stores have a high turnover rate and can be depleted quickly and frequently so that signals exist to monitor and correct for carbohydrate imbalances. Fat stores, on the other hand, are nearly limitless, and turnover is slow and infrequent. Thus, as a rule, increased energy intake results in increased fatness (Bray 1993a; Horton et al. 1995).

Energy Expenditure

Energy expenditure has a number of components, the most important of which is basal metabolic rate (BMR). The total energy cost of any given activity is equal to the BMR plus the work done and the heat produced. For sedentary populations, the BMR may comprise 50 percent to 70 percent of the daily energy expenditure. Wide variations in BMR are not accounted for by food intake, meaning that those people with the highest calculated BMRs are not those who eat the most food. However, BMR is depressed in starvation, in individuals with a restricted caloric intake, and in many obese individuals (Bouchard 1994).

Consistent with these findings are studies of energy expenditure involving both involuntary (for example, fidgeting) and voluntary movement that indicate that physical activity is reduced in obese individuals. However, even though physical activity is reduced, there is a higher cost for activity in overweight individuals, resulting in a tendency toward normal or even high levels of energy expenditure for a particular physical activity. In these instances much of the energy is lost as heat, rather than in muscular work. One problem of cross-sectional data is that finding an average level of energy intake for an individual who is already obese tells us little about previous levels of energy intake and energy expenditure that may have contributed to the development of obesity in the first place.

It is likely that genetic predisposition to obesity lies not only in a lower BMR but also in the reduction of heat production that occurs following a meal, that is, in lower diet-induced thermogenesis. A subnormal thermogenic response to food has been reported in clinical experiments among obese individuals and those who have been obese in the past (the post-obese), as well as among subjects maintaining a desirable weight on relatively low food intake. In sum, obese individuals are metabolically more efficient than lean individuals.

Taste Preference and Obesity

Taste preferences represent a major determinant of food intake and have been linked to obesity and weight gain. Genetic differences in taste preferences are heritable (Perusse and Bouchard 1994). For example, sensitivity to phenylthiocarbamide (PTC) is controlled by a major single gene. Individuals who can taste PTC, a bitter synthetic compound, avoid or

reduce intake of foods containing chemically similar, but naturally occurring, compounds, such as cabbage, broccoli, and Brussels sprouts. Research indicates that some obese individuals may have an elevated preference for foods high in carbohydrates and fats, such as ice cream, chocolate, or pastries (Drewnowski 1988). In contrast, anorectic subjects show a preference for sweet but not for fatty foods. Food preferences may be regulated, in part, by peptides that stimulate the central nervous system, increasing neurotransmitters, such as beta endorphins, and, consequently, feelings of well-being. Finally, simply tasting (not swallowing) fat, either in cream cheese or peanut butter, increases insulin production and serum triglycerides. Such findings suggest that sensory receptors in the mouth initiate digestive responses not triggered directly by the nutrient and that the sensory qualities of fat may promote preference for fatty foods (Radloff 1996).

Psychological Factors

In recent years there has been a major change in discussions concerning the psychological aspects of obesity. In earlier psychogenic theories, obese persons were assumed to suffer from emotional disturbances and failures of impulse control. However, systematic assessment of the nature and extent of psychological problems of obese individuals have changed the psychogenic views of obesity to a somatogenic one. The psychosocial problems in question arise primarily from the stigma attached to obesity in contemporary societies.

Genetic Factors

Investigations of the genetic factors in human obesity have developed rapidly. Traditional family studies of parent, offspring, and sibling data identify the extent to which obesity is familial, but they pose the problem of separating the shared environment of families from genetic factors. Adoption studies overcome some of the difficulties encountered in family studies because individuals have a shared environment but not a biological relationship. Moreover, identical and fraternal twin studies can shed light on genetic as well as environmental contributions (Bouchard et al. 1988, 1990; Bouchard 1994).

Inheritable features include metabolic rates, hormone and enzyme levels, and the amount and patterning of subcutaneous fat. Data from a number of studies indicate that Body Mass Index has a heritability (range 0 to 1) from 0.4 to 0.6, suggesting that genes may be responsible for approximately one-half of the total phenotypic variation in obesity. Adoption studies found that BMIs correlated more strongly with biological than adoptive parents. Investigations in Denmark and Iowa of adult twins showed a high heritability of 0.8. Additive and nonadditive genetic components point to many obesity-promoting genes (Bouchard 1994).

A number of important genes that control eating and

weight gain have been discovered in mice. These have human analogs, but obesity-promoting mutations of these genes have not been located in humans. Recently, mutations in the leptin receptor gene have been found among the Pimas of Arizona, the Finns, and the French (Bouchard and Bray 1996; Gibbs 1996). The obese gene encodes for leptin, a hormone produced by fat cells. Mice with a mutation in this gene produce either no leptin or a malformed version and are obese with weights up to three times those of normal mice. The diabetes gene codes for a receptor protein that responds to leptin by reducing appetite and increasing metabolism. Mice with a mutation of this gene do not receive the leptin signal and get very fat from infancy. Other genes are “fat” and “tubby.” Mice with mutations in either gene put on weight gradually, more like the human pattern. Recent work on leptin and leptin receptors has focused on developing appropriate pharmacological interventions. Leptin causes obese and normal mice to lose weight by signaling the brain that the body has enough fat, which in turn suppresses appetite.

The search for genes and their products that control appetite led to the discovery in 1996 of the hormone urocortin, which is a powerful appetite suppressant in rats. Other promising research has concentrated on a neuropeptide that stimulates appetite and on lipoprotein lipase, an enzyme involved in fat deposition (Gibbs 1996).

Consequences of Obesity

Mortality

A number of studies have focused on mortality and morbidity associated with overweight and obesity. Those done in Norway, Canada, and the United States indicate that BMIs associated with lowest mortality lie in the range of 15 percent below to 5 percent above ideal weight, although recently, some physicians have argued that failure to control for the effects of smoking produces an artificially high mortality in leaner subjects. All studies with more than 20,000 participants have shown a positive relationship between overweight and mortality. But the Build Study of 1979 indicated that above-average weights are associated with optimal life expectancy. Moreover, 40 percent of all smaller employee, community, and random population studies have failed to demonstrate a relationship between body weight and mortality (Sjostrom 1993).

Similarly, the First National Health and Nutrition Examination Survey 9-year follow-up found that there was no additional risk associated with overweight among women and a statistically significant but moderate additional risk (relative risk 1.1 to 1.2) for men ages 55 to 74 years. Low body weight, however, was associated with increased mortality (relative risk 1.3 to 1.6), except for women age 55 to 64 years. These results suggest a need for clinically specific definitions of obesity and overweight, especially among the elderly (Tayback, Kumanyika, and Chee 1990).

A change in weight has also been associated with mortality in adulthood. The Nurses' Study, using age 18 as the standard, showed that increases in weight up to 9.9 kg either reduced or left unchanged the relative risk of mortality. However, weight increases of 10 to 19.9 kg and 20 to 34.9 kg resulted in relative risks of 1.7 and 2.5, respectively. In the Framingham Study, an estimated 10 percent loss in body weight corresponded with a 20 percent reduction in the risk of developing coronary artery disease (Committee on Diet and Health 1989). Independent of weight or BMI, a high waist-to-hip ratio or large deposits of fat in the center of the body (related to internal stores of visceral fat) are associated with increases in mortality and morbidity.

Data also suggest that obesity as assessed by BMI or other measures may mean different mortality and morbidity risks for different populations. For example, overall mortality was not associated with obesity among the Pima except for the most obese men (Knowler et al. 1991). Among Japanese-American men in the Honolulu Heart Program, the BMI, subscapular skinfold thickness and central obesity predicted coronary artery disease, but only subscapular skinfold thickness predicted stroke (Curb and Marcus 1991b).

Morbidity

A wide variety of disorders and problems are caused or exacerbated by obesity (National Institutes of Health 1985; Bouchard and Johnston 1988), and T. Van Itallie (1985) has listed these by organ system. For the cardiovascular system they are premature coronary artery disease, ventricular arrhythmia and congestive heart failure, hypertension, stroke, and varicose veins. Obesity can affect the respiratory system in terms of alveolar hypoventilation or the Pickwickian syndrome, obstructive sleep apnea, and ventricular hypertrophy. Under the digestive system rubric there may be an increase in gall bladder and liver diseases. For the hormonal and metabolic systems, diabetes mellitus, gout, and hyperlipidemia are the most common results of obesity. Kidneys may be affected, resulting in proteinuria and renal vein thrombosis. The skin may develop striae and plantar callus, and osteoarthritis of the knee and spine are exacerbated by obesity. Obesity increases the risk of endometriosis and breast cancer in women and can impair reproductive and sexual functions. Obesity also enhances the risk of surgical and anesthetic procedures and reduces physical agility, which can lead to accident proneness. Finally, obesity may interfere with the diagnosis of other disorders by physically obscuring their presence.

Social Stigma of Obesity

Obese individuals in the United States suffer from social and psychological prejudice. Even children hold these prejudices. When a group of 6-year-olds were shown a fat person's silhouette and asked to describe

the person's characteristics, they said "lazy, cheating and lying" (Czajka-Narins and Parham 1990). In another study where children were shown drawings of an obese child, a child in a wheelchair, a child on crutches, a facially disfigured child, and a child amputee, they disliked only the latter more than the overweight child. Furthermore, children prefer to play with thin rather than fat rag dolls, and parents prefer to have thin rather than obese children photographed.

Obese adolescent females have reported fewer dates and less participation in school organizations than nonobese adolescents. College students, when asked whom they were least likely to marry, ranked obese individuals fifth lowest in desirability, following an embezzler, a cocaine user, an ex-mental patient, and a shoplifter. Adults rate nonobese figures as happier, having more friends, smarter, more attractive, less lonely, and less mean than obese persons. College students in another study stated that obese individuals were warm and friendly but also unhappy, without self-confidence, self-indulgent, undisciplined, lazy, and unattractive. It is interesting, however, that psychological profiles of obese individuals show personality characteristics and achievement levels that belie these prejudicial views (Probart and Lieberman 1992).

Health professionals in the United States also hold negative views of obese individuals. They consider obese individuals to be hypochondriacal, possessing impaired judgment, having inadequate hygiene, and indulging in inappropriate and self-injurious behavior. Additionally, they find obesity to be aesthetically displeasing. As mentioned, there is an active segment of the U.S. population working to counter such widespread prejudices (Price et al. 1987).

Treatment for Obesity

Dieting

Dieting is a routine aspect of life for many Americans, even among some who are not obese. Studies indicate that in the United States, 33 to 44 percent of adult women and 22 to 34 percent of adult men are actively dieting at any given time and that Americans spend an estimated 30 to 50 billion dollars a year on weight-loss programs. The incidence of dieting varies little by ethnicity for women, but among men, Hispanic Americans have the highest proportion of dieting and African-Americans the lowest. Dieting is most common among well-educated and higher socioeconomic classes (Kopelman et al. 1994; Elmer-Dewitt 1995).

Reports of dieting by high-school-age students are common. The National Adolescent Student Health Survey of eighth and tenth graders revealed that 61 percent of girls and 28 percent of boys reported dieting to lose weight. However, a longitudinal study of Arizona adolescents by the Nichters (1991) found extensive discourse about dieting but little modification of

food behaviors or weight loss. Very few girls in this study reported using vomiting, diet pills, laxatives, or diuretics as weight-control methods. Many girls responded that they did not "diet to lose weight" but rather "watched what they ate continually" and avoided "fattening foods." Eighty-five percent rarely or never counted calories, nor did they know how many calories they consumed.

Although caloric restriction (usually 1,200 to 1,000 kcal/day) should be the only method of treatment for obesity, long-term analyses of results show that 95 percent of those who are successful at weight loss regain it within one to two years. Very low calorie diets (400 to 800 kcal/day) lead to rapid weight loss, but there is little evidence that long-term weight maintenance is improved. Repeated weight gain and loss cycles are called weight cycling or yo-yo dieting. Although there are questions about altered metabolism with weight cycling, lean tissue mass does not appear to be lost at an increased rate with each cycle (Garner and Wooley 1991).

Many weight-loss programs provide special foods low in calories and high in fiber as part of a comprehensive program. Weight-reduction clinics and lay support groups use a combination of diet with behavioral and exercise intervention. These approaches have costs associated with them, ranging from expensive diet clinics that provide specialized care with a multidisciplinary team of psychologists, dietitians, physicians, and exercise specialists to inexpensive lay support groups, such as Weight Watchers, Take Off Pounds Sensibly (TOPS), and Overeaters Anonymous. Recently, work-site and school weight-control groups have also been formed, emphasizing both weight loss and relapse prevention.

Behavioral therapy as an adjunct to dietary treatment improves compliance and may improve long-term results. Elements of behavioral therapy include self-monitoring of dietary intake; control of external eating stimuli; analyses and changes in eating behavior, including speed, time of day, and locus of activity; rewards for weight loss and weight maintenance; nutritional education; physical activity to improve both weight loss and overall well-being; and cognitive restructuring, concentrating on positive goals to counter negative, self-defeating thoughts (Kanarek et al. 1984).

Exercise

A number of mechanisms have been proposed to explain the association between exercise and weight control. The most frequently advanced benefits are an increase in lean body mass, resulting in a higher basal metabolic rate; increases in metabolic rate produced by the exercise and enduring beyond specific bouts of exercise; the energy expenditure of the activity itself; and the psychological benefits, including self-esteem, modulation of mood, and improved body image. The current recommendation is for regular

low-level exercises (a heart rate below 60 or 70 percent of the maximum) that are beneficial and easy to maintain over a long duration.

A number of techniques can be employed to improve adherence to an exercise plan. These include setting goals, a program of relapse prevention, and behavior changes involving stimulus control and reinforcement in a social environment, perhaps including one's spouse. When they begin to exercise, many overweight individuals face a combination of physical and psychosocial burdens related to negative feelings associated with past experience. Some studies indicate that exercise alone can produce a modest gain in lean body mass and a loss in fat in weight-stable individuals. However, both animal and human experiments show that exercise does not conserve lean weight in the face of significant energy deficit (Stunkard and Wadden 1992; Bouchard and Bray 1996).

Drugs

In cases of extreme obesity or when there is fear of comorbidities associated with a central distribution of fat, drug therapy may be warranted. These drugs include agents that act on the noradrenergic and serotonergic systems, opioid receptors, and peptide agonists or antagonists. The most popular drugs – fenfluramine, fluoxetine, or dexfenfluramine and phentermine – increase serotonin levels. Other classes of drugs include thermogenic drugs, growth hormone agonists, and drugs that act directly on the gastrointestinal system, including enzyme inhibitors and inhibitors of absorption. These drugs are not without risk (heart valve abnormalities and pulmonary hypertension, for example, have been linked to a phentermine–fenfluramine pill) and, in fact, have not been used with substantially large populations long enough to assess long-term benefits or risks. This is partly because there was a hiatus of about 20 years in the United States when drugs were not developed specifically to treat obesity. Now, however, there are a number of pharmaceutical companies in the process of developing such new drugs (Gibbs 1996).

Surgery

A surgical procedure may be warranted for morbidly obese individuals. This may be in the form of liposuction or surgical reduction of fat deposits, usually from the abdominal apron. But food ingestion can be restricted by temporarily wiring the jaws, or its absorption can be reduced either through the reduction of stomach capacity by stapling or by creating a small intestinal bypass. Surgical intervention, however, is not without risk and sequelae, such as severe diarrhea.

Summary

Overweight and overfatness (that is, obesity) are conditions of epidemic proportions related to food consumption and activity patterns. The human ability to

store fat and selectively mobilize it evolved as a defense against food shortages, cold climates, demanding physical labor, disease, and the physiological requirements of pregnancy and lactation. Our understanding of the physiological and genetic basis for obesity has increased considerably in the last decade.

Obesity has been linked to increased mortality and morbidity in nearly all organ systems. Moreover, psychosocial stigmatizing of the obese, at least in affluent countries, has an adverse impact on their education, occupation, social interaction, and self-esteem.

In the United States, a multibillion-dollar “diet industry” has developed that includes new diet products, medications, clinics, health clubs, support groups, and obesity specialists. With all of these efforts, including the increasing availability of fat- and sugar-free foods, we continue to get fatter – in the United States and around the world.

Leslie Sue Lieberman

Bibliography

- Abraham, S., M. D. Carroll, M. F. Najjar, and R. Fulwood. 1983. *Obese and overweight adults in the United States*. Hyattsville, Md.
- Abraham, S., and C. L. Johnson. 1980. Prevalence of severe obesity in adults in the United States. *American Journal of Clinical Nutrition* 33: 364–9.
- Abraham, S., C. L. Johnson, and M. F. Najjar 1979. *Weight and height of adults 18–74 years of age: United States 1971–7: Age, sex, and race distribution for weight and height*. Hyattsville, Md.
- Ackerknecht, Erwin H. 1982. *A short history of medicine*. Rev. edition. Baltimore, Md.
- Allon, N. 1982. The stigma of obesity in everyday life. In *Psychological aspects of obesity: A handbook*, ed. B. Wolman, 130–74. New York.
- Aluli, Non E. 1991. Prevalence of obesity in a native Hawaiian population. *American Journal of Clinical Nutrition* 53: 1556S–60S.
- Ashwell, M. 1994. Obesity in men and women. *International Journal of Obesity* 18 (Supplement 1): 51–7.
- Associated Press. 1996. Many Americans are overweight poll finds. *Gainesville Sun* (Fla.), February 5, p. 3A.
- Atkinson, R. L., and V. S. Hubbard. 1994. Report on the National Institute of Health workshop on pharmacologic treatment of obesity. *American Journal of Clinical Nutrition* 60: 153–6.
- Baba, S., and P. Zimmet, eds. 1990. *World data book on obesity*. New York.
- Banting, W. 1864. *Letter on corpulence addressed to the public*. Third edition. London.
- Beller, A. S. 1977. *Fat and thin: A natural history of obesity*. New York.
- Bindon, J. R. 1995. Polynesian responses to modernization: Overweight and obesity in the South Pacific. In *Social aspects of obesity*, ed. I. de Garine and N. J. Pollock, 227–51. Amsterdam.
- Bindon, J. R., and P. T. Baker. 1985. Modernisation, migration and obesity among Samoan adults. *Annals of Human Biology* 12: 67–76.
- Bolton-Smith, C., and M. Woodward. 1994. Dietary composi-

- tion and fat to sugar ratios in relation to obesity. *International Journal of Obesity* 18: 820-8.
- Bouchard, C. 1994. *The genetics of obesity*. Boca Raton, Fla.
- Bouchard, C., and G. A. Bray. 1996. *Regulation of body weight: Biological and behavioral mechanisms*. New York.
- Bouchard, C., and F. E. Johnston. 1988. *Fat distribution during growth and later health outcomes*. New York.
- Bouchard, C., A. Tremblay, J. P. Despres, et al. 1988. Sensitivity to overfeeding: The Quebec experiment with identical twins. *Progress in Food and Nutrition Science* 12: 45-72.
1990. The response to long-term overfeeding in twins. *New England Journal of Medicine* 322: 1477-90.
- Bradden, F. E. M., B. Rogers, M. E. J. Wadsworth, and J. M. C. Davis. 1988. Onset of obesity in a 36 year birth cohort study. *British Medical Journal* 273: 297-303.
- Bray, G. A. 1987. Obesity - A disease of nutrient or energy balance? *Nutrition Review* 45: 33-43.
1988. Perspective: Future research in obesity. In *Diet and obesity*, eds. G. A. Bray, J. LeBlanc, S. Inoue, and M. Suzuki, 229-43. Tokyo.
1992. Obesity: Historical development of scientific and cultural ideas. In *Obesity*, ed. P. Björntorp and B. N. Brodoff, 281-94. Philadelphia, Pa.
- 1993a. The nutrient balance approach to obesity. *Nutrition Today* 28: 13-18.
- 1993b. *Obesity: Comparative methods of weight control*. Westport, Conn.
- Bray, G. A., ed. 1979. *Obesity in America*. HEW (PHS) National Institute of Health Publication No. 79-359. Washington, D.C.
- Brink, P. J. 1995. Fertility and fat: The Annang fattening room. In *Social aspects of obesity*, ed. I. de Garine and N. J. Pollock, 71-85. Amsterdam.
- Broussard, B. A., A. Johnson, J. Himes, et al. 1991. Prevalence of obesity in American Indians and Alaskan natives. *American Journal of Clinical Nutrition* 53: 1535S-1542S.
- Brown, P. J. 1991. Culture and the evolution of obesity. *Human Nature* 2: 31-57.
1993. Cultural perspectives in the etiology and treatment of obesity. In *Obesity: Theory and therapy*. Second edition, ed. J. J. Stunkard and T. A. Wadden, 179-93. New York.
- Brown, P. J., and M. Konner. 1987. An anthropological perspective on obesity. *Annals of the New York Academy of Science* 499: 29-46.
- Brownell, K. D. 1995. Exercise in the treatment of obesity. In *Eating disorders and obesity: A comprehensive handbook*, ed. K. D. Brownell and C. G. Fairburn, 473-8. New York.
- Burch, H. 1973. *Eating disorders: Obesity, anorexia, and the person within*. New York.
- Burwell, C. S. 1956. Extreme obesity associated with alveolar hypoventilation - a Pickwickian syndrome. *American Journal of Medicine* 21: 811-18.
- Butler, M. G. 1990. Prader-Willi syndrome: Current understanding of cause and diagnosis. *American Journal of Human Genetics* 35: 319-32.
- Cassell, J. A. 1995. Social anthropology and nutrition: A different look at obesity in America. *Journal of the American Diabetic Association* 95: 424-7.
- Cassidy, C. 1991. The good body: When big is better. *Medical Anthropology* 13: 181-213.
- Cavalli-Sforza, L. L. 1981. Human evolution and nutrition. In *Food, nutrition and evolution: Food as an environmental factor in the genesis of human variability*, ed. D. N. Walcher and N. Kretchmer, 1-7. New York.
- Committee on Diet and Health, National Research Council. 1989. *Diet and health: Implications for reducing chronic disease risk*. Washington, D.C.
- Conway, J. 1995. Ethnicity and energy stores. *American Journal of Clinical Nutrition* 52: 1067S-71S.
- Crawford, M., and D. Marsh. 1989. *The driving force: Food, evolution and the future*. London.
- Curb, J. D., and E. B. Marcus. 1991a. Body fat in Japanese Americans. *American Journal of Clinical Nutrition* 53: 1552S-5S.
- 1991b. Body fat, coronary heart disease and stroke in Japanese men. *American Journal of Clinical Nutrition* 53: 1612S-15S.
- Czajka-Narins, D. M., and E. S. Parham. 1990. Fear of fat: Attitudes toward obesity. *Nutrition Today* 25: 26-32.
- Dawson, D. 1988. Ethnic differences in female overweight: Data from the 1985 National Health Interview Study. *American Journal of Public Health* 78: 1326-9.
- de Garine, I., and N. J. Pollock, eds. 1995. *Social aspects of obesity*. Amsterdam.
- Dietz, W. H., and S. L. Gortmaker. 1985. Do we fatten our children at the television set? Obesity and television viewing in children and adolescents. *Pediatrics* 75: 807-12.
- Drewnowski, A. 1988. Obesity and taste preferences for sweetness and fat. In *Diet and obesity*, ed. G. A. Bray, J. LeBlanc, S. Inoue, and M. Suzuki, 101-12. Tokyo.
- Dullo, A. G., and D. S. Miller. 1987. Obesity: A disorder of the sympathetic nervous system. *World Review of Nutrition and Diet* 50: 1-56.
- Eaton, S. B., M. Shostak, and M. Konner. 1988. *The paleolithic prescription: A program of diet and exercise and a design for living*. New York.
- Edholm, O. G. 1973. Energy expenditure and food intake. In *Energy balance in man*, ed. M. Apfelbaum, 51-60. Paris.
- Elmer-Dewitt, P. 1995. Fat times. *Time*, January 16, pp. 58-65.
- Epstein, F. H., and M. Higgins. 1952. Epidemiology of obesity. In *Obesity*, ed. P. Björntorp and B. Brodoff, 330-42. Philadelphia, Pa.
- Folsom, A. R., G. L. Burke, C. L. Byers, et al. 1991. Implications of obesity for cardiovascular disease in blacks: The CARDIA and ARIC studies. *American Journal of Clinical Nutrition* 53: 1604S-11S.
- Forbes, G. B. 1992. Exercise and lean weight: The influence of body weight. *Nutrition Reviews* 50: 157-61.
- Garfinkel, L. 1985. Overweight and cancer. *Annals of Internal Medicine* 103: 1034-6.
- Garn, S. M., P. E. Cole, and S. M. Bailey. 1979. Living together as a factor in family-line resemblances. *Human Biology* 51: 565-87.
- Garner, D. M., and S. C. Wooley. 1991. Confronting the failure of behavioral and dietary treatment for obesity. *Clinical Psychological Review* 11: 729-80.
- Gibbs, W. W. 1996. Gaining on fat. *Scientific American* 275: 88-94.
- Goldblatt, P. B., M. E. Moore, and A. J. Stunkard. 1965. Social factors in obesity. *Journal of the American Medical Association* 192: 1039-44.
- Goodman, N., S. A. Richardson, S. M. Dombusch, and A. H. Hastorf. 1963. Variant reactions to physical disabilities. *American Sociological Review* 28: 429-35.
- Gotmaker, S. L., A. Must, and J. M. Perrin. 1993. Social and economic consequences of overweight in adolescence and young adulthood. *New England Journal of Medicine* 329: 1008-19.
- Gurney, M., and J. Gorstein. 1988. The global relevance of obesity: An initial overview of available data. *World Health Statistical Quarterly* 41: 251-4.

- Harlan, W. R., J. R. Landis, K. M. Flegal, et al. 1988. Secular trends in body mass in the United States 1960–1980. *American Journal of Epidemiology* 128: 1065–74.
- Harris, M. B., and S. D. Smith. 1983. The relationships of age, sex, ethnicity and weight to stereotypes of obesity and self-perception. *International Journal of Obesity* 7: 361–71.
- Hazuda, H. P., B. Mitchell, S. Hoffner, and M. Stern. 1991. Obesity in Mexican American subgroup: Findings from the San Antonio Heart Study. *American Journal of Clinical Nutrition* 53: 1529S–34S.
- Himes, J. H. 1991. *Anthropometric assessment of nutritional status*. New York.
- Hippocrates. 1849. *The genuine works of Hippocrates translated from the Greek with a preliminary discourse and annotations*, trans. F. Adams. London.
- Hofstetter, A., Y. Schutz, E. Jequier, and J. Wahren. 1986. Increased 24-hour energy expenditure in cigarette smokers. *New England Journal of Medicine* 314: 79–82.
- Horton, T., H. Drougas, A. Brachey, et al. 1995. Fat and carbohydrate overfeeding in humans: Different effects on energy storage. *American Journal of Clinical Nutrition* 62: 19–29.
- Hurley, J., and G. Collins. 1997. Let use be your guide. *Nutrition Action Health Letter* 24: 8–9.
- Institute of Medicine. 1995. *Weighing the options: Criteria for evaluating weight management programs*. Washington, D.C.
- Kahn, H., and D. Williamson. 1991. Is race associated with weight change in U.S. adults after adjustment for income, education and marital factors? *American Journal of Clinical Nutrition* 53: 1566S–70S.
- Kanarek, R. B., N. Orthen-Gambill, R. Marks-Kaufman, and J. Mayer. 1984. Obesity: Possible psychological and metabolic determinants. In *Human Nutrition: A Comprehensive Treatise*. No. 5 *Nutrition and Behavior*, ed. J. R. Geller, 339–96. New York.
- Knowler, W. C., D. J. Pettitt, M. Sood, et al. 1991. Obesity in the Pima Indians: Its magnitude and relationship with diabetes. *American Journal of Clinical Nutrition* 53: 1543S–51S.
- Kopelman, P. G., N. Finer, K. R. Fox, et al. 1994. Association for the Study of Obesity (ASO) consensus statement on obesity. *International Journal on Obesity* 18: 189–91.
- Kuczmarski, R. J., K. M. Flegal, S. M. Campbell, and C. L. Johnson. 1994. Increasing prevalence of overweight among U.S. adults. The National Health and Nutrition Examination Surveys, 1960–1991. *Journal of American Medical Association* 272: 205–11.
- Kumanyika, S. 1993. Special issues regarding obesity in minority populations. *Annals of Internal Medicine* 119: 550–4.
1994. Obesity in minority populations: An epidemiological assessment. *Obesity Research* 2: 166–82.
- Lachance, P. A. 1994. Scientific status summary: Human obesity. *Food Technology* (February): 127–36.
- Leibel, R. L., N. K. Edens, and S. K. Fried. 1989. Physiologic basis for the control of body fat distribution in humans. *Annual Review of Nutrition* 9: 417–43.
- Leigh, J. P., J. Fries, and H. B. Hubert. 1992. Gender and race differences in correlation between body mass and education. *Journal of Epidemiology and Community Health* 46: 191–6.
- Levitsky, D. A., and R. Troiano. 1992. Metabolic consequences of fenfluramine for the control of body weight. *American Journal of Clinical Nutrition* 55: 167S–72S.
- Lieberman, L. S., and C. K. Probart. 1992. Body weight: The ideal, the normal and the desirable. *Collegium Antropologicum* 16: 279–95.
- Lohman, T. G. 1992. *Advances in body composition assessment*. Champaign, Ill.
- Low-cal is booming all over the market. *Progressive Grocer* 49: 1169.
- Luskas, H. C. 1987. Methods for the assessment of human body composition: traditional and new. *American Journal of Clinical Nutrition* 46: 537–58.
- Martin, R. J., B. D. White, and M. G. Hulsey. 1991. The regulation of body weight. *American Scientist* 79: 528–41.
- Massara, E. B. 1989. *Que gordita! A study of weight among women in a Puerto Rican community*. New York.
- Mayer, J. 1969 *Overweight: Causes, costs and controls*. Englewood Cliffs, N.J.
- McGarvey, S. 1991. Obesity in Samoans and a perspective on its etiology in Polynesians. *American Journal of Clinical Nutrition* 53: 1586S–94S.
- McMurphy, M. P., M. T. Sesqueria, S. L. Conner, and W. E. Conner. 1991. Changes in lipid and lipoprotein levels and body weight in Tarahumara Indians after consumption of an affluent diet. *New England Journal of Medicine* 325: 1704–8.
- Melnik, M. G., and E. Weinstein. 1994. Preventing obesity in black women by targeting adolescents: A literature review. *Journal of the American Dietetic Association* 94: 536–40.
- Metropolitan Life Insurance Company. 1983. Metropolitan height and weight tables. *Statistical Bulletin of the Metropolitan Insurance Company* 64: 2–9.
- Meyer, J., and A. J. Stunkard. 1993. Genetics and human obesity. In *Obesity: Theory and therapy*. Second edition, ed. A. J. Stunkard and T. A. Wadden. New York.
- Najjar, M. F., and M. Rowland. 1987. *Anthropometric reference data and prevalence of overweight: United States 1976–1980*. Washington, D.C.
- National Institutes of Health. 1985. Health implications of obesity: National Institute of Health consensus development conference statement. *Annals of Internal Medicine* 103: 1073–7.
1992. *Weight loss and obesity*. Bethesda, Md.
- Nichter, N., and M. Nichter. 1991. Hype and weight. *Medical Anthropology* 13: 249–84.
- O'Dea, K. 1984. Marked improvement in carbohydrate and lipid metabolism in diabetic Australian Aborigines after temporary reversion to traditional lifestyle. *Diabetes* 33: 596–603.
- Perri, M. C., S. F. Sears, and J. E. Clark. 1993. Strategies for improving maintenance of weight loss: Toward a continuous care model of obesity management. *Diabetes Care* 16: 200–9.
- Perusse, L., and C. Bouchard. 1994. Genetics of energy intake and food preferences. In *The genetics of obesity*, ed. C. Bourhard. Boca Raton, Fla.
- Pollock, N. J. 1995. Social fattening patterns in the Pacific: The positive side of obesity. In *Social aspects of obesity*, ed. I. de Garine and N. J. Pollock, 87–109. Amsterdam.
- Price, J. H., S. M. Desmond, R. A. Krol, et al. 1987. Family practice physicians' beliefs, attitudes and practices regarding obesity. *American Journal of Preventive Medicine* 3: 339–45.
- Probart, C. K., and L. S. Lieberman. 1992. Cultural influences on normal and idealized female body size. *Collegium Antropologicum* 16: 297–301.
- Radloff, J. 1996. The taste of fat may pose a heart risk. *Science News* 149: 373.

- Ravelli, G. P., Z. A. Stein, and M. W. Susser. 1976. Obesity in young men after famine exposure in utero and early infancy. *New England Journal of Medicine* 295: 347-53.
- Ravussin, E., M. Valencia, J. Esparza, et al. 1994. Effects of a traditional lifestyle on obesity in Pima Indians. *Diabetes Care* 17: 1067-74.
- Ritenbaugh, C. 1982. Obesity as culture-bound syndrome. *Cultural, Medicine and Psychiatry* 6: 347-61.
1991. Body size and shape: A dialogue of culture and biology. *Medical Anthropology* 13: 173-80.
- Roberts, D. F. 1953. Body weight, race and climate. *American Journal of Physical Anthropology* 11: 533-58.
- Rodin, J. 1981. Current status of the internal-external hypothesis for obesity. *American Psychologist* 36: 361-92.
- Rolls, B. J. 1991. Effects of artificial sweeteners on hunger, food intake and body weight: A review. *American Journal of Clinical Nutrition* 53: 872-8.
- Schacter, S., and J. Rodin. 1974. *Obese humans and rats*. Potomac, Md.
- Schlundt, O. G., D. Taylor, J. O. Hill, et al. 1991. A behavioral taxonomy of obese female participants in a weight loss program. *American Journal of Clinical Nutrition* 53: 1151-8.
- Schwartz, H. 1986. *Never satisfied: A cultural history of diet, fantasies and fat*. New York.
- Silverstone, J. T. 1993. Appetite-suppressant drugs. In *Obesity: Theory and therapy*. Second edition, ed. A. J. Stunkard and T. A. Wadden, 275-85. New York.
- Sims, E. A. H., and E. S. Horton. 1968. Endocrine and metabolic adaptation to obesity and starvation. *American Journal of Clinical Nutrition* 21: 1455-70.
- Sjostrom, L. 1993. Impacts of body weight, body composition and adipose tissue distribution on morbidity and mortality. *Obesity: Theory and therapy*. Second edition, ed. A. J. Stunkard and T. A. Wadden, 13-41. New York.
- Sobal, J., and A. J. Stunkard. 1989. Socioeconomic status and obesity: A review of the literature. *Psychological Bulletin* 105: 260-75.
- Society of Actuaries. 1959. *Build and blood pressure study*. Washington, D.C.
- Society of Actuaries and Association of Life Insurance Medical Directors of America. 1979. *Build Study*. Chicago.
- Stern, M. P., C. Gonzales, D. B. Mitchell, et al. 1992. Genetics and environmental determinants of type II diabetes in Mexico City and San Antonio. *Diabetes* 41: 484-92.
- Stevens, J., J. E. Keil, P. F. Rust, et al. 1992. Body mass index and body girths as predictors of mortality in black and white women. *Archives of Internal Medicine* 153: 1257-62.
- Stillman, I. M., and S. J. Baker. 1967. *The doctor's quick weight loss diet*. Englewood Cliffs, N.J.
- Stunkard, A. J. 1979. Obesity and social environment: Current status, future prospects. In *Obesity in America*. HEW (PHS) NIH Publication No. 79-359, ed. G. A. Bray. Washington, D.C.
- Stunkard, A. J., A. Y. Cohen, and M. R. Felix. 1989. Weight loss competitors and the worksite: How they work and how well. *Preventive Medicine* 18: 460-74.
- Stunkard, A. J., J. R. Harris, N. L. Pederson, and G. E. McClearn. 1990. The body mass index of twins who have been reared apart. *New England Journal of Medicine* 298: 1483-99.
- Stunkard, A. J. and T. I. A. Sorensen. 1992. Obesity and socioeconomic status - A complex relationship. *New England Journal of Medicine*. 239: 1036-7.
- Stunkard, A., and T. Wadden. 1992. Psychological aspect of human obesity. In *Obesity*, ed. P. Björntorp and B. N. Brodoff, 352-60. Philadelphia, Pa.
- Styles, M. H. 1980. Soul, black women and food. In *A woman's conflict: The special relationship between women and food*, ed. J. R. Kaplan. 161-76. Englewood Cliffs, N.J.
- Sugar consumption and health: Summary of a critical review of the literature. 1995. *Nutrition Update* (Summer): 2-4.
- Suzuki, M., and N. Hosoya. 1988. Obesity and alteration of food habits with westernization. In *Diet and obesity*, eds. G. A. Bray, J. LeBlanc, S. Inoue, and M. Suzuki, 1-8. Tokyo.
- Tayback, M., S. Kumanyika, and E. Chee. 1990. Body weight as a risk factor in the elderly. *Archives of Internal Medicine* 150: 1065-72.
- Troisi, R. J., J. W. Heinold, P. S. Vokonas, and S. T. Weiss. 1991. Cigarette smoking, dietary intake and physical activity: Effects on body fat distribution - The Normative Aging Study. *American Journal of Clinical Nutrition* 53: 1104-11.
- Trowell, H. C., and D. P. Burkitt. 1981. *Western diseases and their emergence and prevention*. Cambridge, Mass.
- Tucker, L., and M. Bagwell. 1991. Television viewing and obesity. *American Journal of Public Health* 81: 908-11.
- Van Itallie, T. B. 1985. Health implications of overweight and obesity in the United States. *Annals of Internal Medicine* 103: 983-8.
1992. Body weight, morbidity and longevity. In *Obesity*, ed. P. Björntorp and B. N. Brodoff, 361-9. Philadelphia, Pa.
- Wadden, T. A. 1993. The treatment of obesity. In *Obesity: Theory and therapy*. Second edition, ed. A. J. Stunkard and T. A. Wadden, 197-217. New York.
- Wadden, T. A., and A. J. Stunkard. 1987. Psychopathology and obesity. *Annals of the New York Academy of Science* 499: 55-65.
- Wardd, W. 1829. *Comments on corpulency, liniments and leanness*. London.
- Weigley, E. S. 1984. Average? Ideal? Desirable? A brief overview of height-weight tables in the United States. *Journal of the American Dietetic Association* 84: 417-23.
- Welty, T. K. 1991. Health implications of obesity in American Indians and Alaskan natives. *American Journal of Clinical Nutrition* 53: 1616S-20S.
1994. Fitness through the ages: Stay in shape in your 20s, 30s, 40s and beyond. *American Health* 13: 74.
- White, A., G. Nicolas, K. Foster, et al. 1993. *Health survey for England 1991*. London.
- Wilmsen, E. 1978. Seasonal effects of dietary intake in the Kalahari San. *Federation Proceedings* 37: 65-71.
- World Health Organization (WHO). 1988. *Measuring obesity: Classification and description of anthropometric data*. Report on a WHO consultation on the epidemiology of obesity. Copenhagen.
- Wu, C. 1996. No-cal powder may sub for food's fat. *Science News* 150: 132.



IV.F

Diet and Chronic Disease

IV.F1 Diabetes

Diabetes mellitus (DM) is a heterogeneous group of endocrine disorders characterized by hyperglycemia (high blood sugar levels) during fasting or following a meal. Other characteristic symptoms of diabetes include excessive urination, urine containing sugar, hunger, thirst, fatigue, and weight loss. The disorder is caused by a resistance to the action of insulin, or a lack or insufficient production of insulin to transport glucose from the blood into cells where it is used as the primary energy source for cellular metabolism. Although diabetes has been a recognized disease for at least two millennia, only since the mid-1970s has there been a consensus on the classification and diagnosis of DM.

Insulin-dependent diabetes mellitus, also called juvenile diabetes or Type I diabetes, is an autoimmune disease that generally affects individuals under the age of 20 and has an acute onset. Noninsulin-dependent diabetes mellitus, Type II, or maturity onset diabetes mellitus, has a complex etiology often associated with obesity and most frequently occurs among individuals over 40 years of age. Ninety to 95 percent of diabetes worldwide is of the latter type. Gestational diabetes appears to be a subset of Type II diabetes, and there are rare genetic syndromes, such as hemochromatosis, drugs, and infections, associated with pancreatic diseases that can cause diabetes. The underlying pathophysiology of Type II diabetes involves the increasing resistance of cells, particularly muscle and adipose (fat) cells, to the transport of glucose across the cell membrane. This resistance or impaired glucose tolerance leads to the classic diagnostic criterion of abnormally high blood sugar concentrations. Glucose also appears in other bodily fluids, particularly urine. The presence of sugar in urine is called glycosuria. This was one of the first signs used in the diagnosis of diabetes. The urine of diabetics tastes sweet and, hence, may attract insects. Such observations were made by early Greek physicians.

This chapter focuses on the relationship of diet and diabetes in historical and cross-cultural perspectives. Three areas are addressed: (1) diet and etiology or causation of diabetes; (2) diet and the pathogenesis of diabetes; and (3) diabetes and diet therapy.

Ancient History through the Middle Ages

Physicians in ancient China, Egypt, Greece, and India wrote accurate clinical descriptions of diabetes. As early as the sixth century B.C., Hindu physicians recognized the clinical symptoms of diabetes and attributed them to dietary indiscretion. The Indian physician Susruta, for example, described the disease as the result of overindulgence in rice, flour, and sugar, which caused the urine to be like an elephant's in quantity. To counter excessive intake, Indian physicians in the *Caraka-Samhita* (c. 123) recommended moderate diets high in carbohydrates and fiber.

Diabetes in Indian Ayurvedic medicine has a complex etiology and pathogenesis. In Ayurveda, the human body is composed of three fundamental elements – *doshas*, *dhatus*, and *malas*. Health is maintained by a balance of these elements, and, consequently, a disturbance of equilibrium causes disease. There are three *doshas* that control various aspects of physiology. Each *dosha* has five divisions, and a disturbance in any one of these can cause diabetes. In addition, the body contains 13 categories of *srotas* (channels of circulation through which pass basic tissue elements), *doshas*, and waste products. Lack of exercise, sleeping during the day (thereby suppressing digestive functions), and excessive fatty food and alcohol consumption can disturb the *srota* involved in carrying the vital substances that comprise fat tissue resulting in diabetes.

More specifically, the etiology of diabetes can be attributed to the excessive consumption of yoghurt, the flesh of animals living in water and marshes, and rice or wheat and starch, especially in refined foods. Excessive worry also plays a role in consumption, sleep, and activity patterns. The treatment for diabetes

in traditional Ayurvedic medicine involves reducing the body fat and inadvertently but effectively regulating the function of the pancreas with the use of *kerela* (bitter gourd). It is recommended in a dose of 30 milliliters twice a day, preferably on an empty stomach. Other botanicals are also used, including the long pepper (*Piper longum*), amalaki (*Embolica officinalis*), turmeric (*Curcuma longa*), Indian pennywort (*Hydrocotyle asiatica*), aconite, "Monk's Hood," "Blue Rocket," "Friar's Cap" (genus *Aconitum*), caltrops, ground burra-nut, puncture vine (*Tribulus terrestris*), and pure honey with hot water. Most vegetables are recommended for individuals with diabetes, but any forms of sugar, rice, potato, banana, cereals, and fruit are to be avoided. Fat is to be ingested in limited quantity.

In Chinese medicine, diabetes is a disease characterized by excessive, sweet urine and described in terms of *chi*. *Chi* is an energy force, or power, existing as a balance between "yin" and "yang." Yin is the force that is feminine, negative, cold, and dark, whereas yang is the force that is masculine, positive, warm, and bright. All substances of heaven and earth are divided into five elements: metal, wood, water, fire, and the earth. The body is a microcosm of these elements with a balance of yin and yang. Balance results in the maintenance of health, whereas a disturbance in this equilibrium produces ill health or disease. The excessive intake of sweets can cause obesity that, in turn, leads to excessive *chi* pushing up in the body and trapping yang in the stomach and large intestine, causing diabetes mellitus.

Although these organs are specifically involved, DM is seen as a whole-body malfunction. Diabetes is divided into three kinds of "thirst" based on location. Upper thirst involves the lung and heart, middle thirst the stomach and spleen, and lower thirst the liver and kidneys. People with a greater yin (female) body composition are the most susceptible to these three types of thirst. Traditional treatment modalities include diets low in carbohydrates and fat, light exercise, herbal medicinals, relaxation, and acupuncture.

Because diabetes mellitus had been relatively rare in Asian countries until the 1990s, current recommendations and treatment modalities often follow the Western biomedical models of etiology, pathogenesis, and therapy. G-Hyon Gang, investigating polytherapeutic approaches to diabetes in contemporary Korea, found that a majority of diabetics employ biomedical treatment plans with insulin or oral hypoglycemic agents and diet regulation. Salts and herbal preparations are used as adjuvants.

The Ebers Papyrus (c. 1550 B.C.) contains a prescription for an antidiabetic diet of wheat germ and okra. Both have been shown to lower blood glucose in animals. Early Arabic physicians described abnormal appetites and the loss of sexual function, but they did not note the sweetness of diabetic urine.

Aretaeus the Cappadocian (A.D. 81-138?) is cred-

ited with coining the word "diabetes." In Greek this means "to run through" or "to siphon." Having observed patients with extreme thirst and frequent and copious urination, he reasoned that their bodies acted like siphons sucking in water at one end and discharging it at the other. In his medical textbook, *The Causes and Signs of Acute and Chronic Disease*, he described diabetic patients as tormented by a burning thirst and thought diabetes was the sequela of an acute disease that attacks the bladder and kidneys. Aretaeus did not note that the urine of diabetics is sweet. He prescribed purgation; a mild diet consisting of sweet wine, cereals, and fruit; and steam baths. He advised a diet high in complex carbohydrates, including cereals, groats, and gruels, but one that also included simple sugars, such as those in fruit juices, dates, and raw quinces. The Greek physician Galen (A.D. 130-210?) also considered DM to be a disease of the kidneys and bladder. For Galen, treatment consisted of overcoming the acidity of the humors, slowing the movement of blood, and cooling overheated kidneys.

From the Renaissance to the 1920s

Diabetes was accorded little attention during the Middle Ages, but starting in the fifteenth century, urine became the focus for experimentation, and the amount and sweetness of urine was used to judge the efficacy of various therapies. Paracelsus (A.D. 1493-1541) evaporated the urine of diabetics and referred to the crystalline residue as "salt." He deduced that this "salt" causes the thirst and the production of excessive urine. He believed this was an imbalance of the humors resulting from a combination of "sulfur" and "salts" in the blood that overflowed into the kidneys, causing inflammation and excessive urinary excretion. He recommended steam baths and "julsip."

In the seventeenth century, Thomas Willis (1621-1675) noted the sweet taste of urine and may have been the first European to describe this classic symptom of DM. He prescribed tincture of antimony, as well as water of quicklime with shavings of saffras. He also recommended anise, licorice, and raisins. Like others before him, Willis considered diabetes a disease of the blood. He thought the etiology related to particles of flesh that were dissolved in the blood, releasing the sweetness, and, therefore, recommended thickening the blood with starch and salt. A contemporary, Thomas Sydenham, proposed that diabetes was caused by the incomplete digestion of what he referred to as chyle of the blood. Other contemporaries proposed the treatment of sugar sickness with the use of bleeding, undernutrition, milk diets, opium therapy, and drinking lime water to neutralize acids.

In 1776, Matthew Dobson of Manchester was able to demonstrate and assay the amount of sugar in

urine by evaporating it and weighing the dried residue. Dobson remarked that this residue looked and tasted like “ground sugar,” and, hence, glycosuria became an important diagnostic tool that continues in use today. In fact, measuring the amount of sugar in the urine permits an assessment of the effectiveness of therapeutic interventions. In 1815, Michel E. Chevreul identified sugar (that is, glucose) in blood and, since the 1980s, blood glucose monitoring has been considered preferable to testing the urine for glucose.

Dobson had proposed a therapeutically rational diet high in meat and fat and low in grains and breads for diabetics. John Rollo, working in the late 1790s, prescribed rancid meat (preferably pork) and milk with the objective of restricting sugar intake and reducing glycosuria. He believed that diabetes was the result of a stomach disorder, which, by causing the improper digestion of food, prevented the proper assimilation of sugar. He noted that the amount of urine depended on the type of food that was eaten: Urinary production increased after the ingestion of vegetables and decreased when diets were high in animal fats and proteins. He advised his patients to consume breakfast and bedtime snacks of 24 ounces of milk, 8 ounces of lime water, and bread and butter. A noontime meal featured pudding made from blood and suet, and the evening meal included old, as opposed to fresh, meat from game. John Rollo’s work was based on extensive metabolic studies of obese diabetic patients in the Greenwich Naval Hospital. His findings shifted the focus from the kidneys to the gastrointestinal tract and provided a scientific basis for therapeutic diets high in fat and protein and low in carbohydrates.

Gross autopsies performed in the 1850s and 1860s revealed no pancreatic abnormalities of diabetics. In 1855, the French physiologist Claude Bernard discovered that the liver secretes glucose from the animal starch or glycogen stored in it, whereas Bernard and Moritz Schiff found that the “destruction” of the pancreas in experimental animals did not result in the onset of diabetes. These findings supported the belief

that a liver disease was the source of diabetes, but attention was redirected to the pancreas in 1889, when Joseph von Mering and Oscar Minkowski demonstrated that the complete removal of the pancreas did cause diabetes in dogs.

In the second half of the nineteenth century, work proceeded on two fronts: the anatomical and physiological elucidation of the pancreas and dietary interventions to control diabetes. In 1869, Paul Langerhans, in his M.D. thesis, described previously unrecognized cells in the rabbit pancreas, without suggesting any function for them. In 1893, Gustave Laguesse named these cells the “islets of Langerhans.” Meanwhile, Minkowski, Laguesse, Eugene Opie, and others demonstrated that these cells produced an internal secretion that controlled glycosuria.

Observations by Apollinaire Bouchardat, culminating in a book on glycosuria in 1875, distinguished between insulin-dependent and noninsulin-dependent diabetics. Bouchardat imposed a very low carbohydrate diet to control glycosuria. Since then, recommendations for the inclusion of carbohydrates in the diet have varied enormously, depending on the prevailing ideas concerning pathophysiology and observations of patients’ success (Table IV.F.1.1). Based on the earlier work of Rollo, meat-based diets were very popular in the middle and late 1800s. Some diets included non-sucrose sugars, such as the sugar alcohols (for example, sorbitol), and milk sugar or lactose. Low-starch vegetables, such as celery and Jerusalem artichokes, were also included.

Von Mering proposed a meal plan of 250 grams (g) of meat, 80 to 120 g of rice, semolina, or buckwheat grits, and an unlimited amount of stale bread. When carbohydrates were advocated as part of dietary control, they were generally high fiber, complex, starchy carbohydrates, such as oats. The starvation regimen often found in the late 1800s and prior to the discovery of insulin in 1921 was, in part, based on Bouchardat’s observation that during the siege of Paris in 1871, the imposed starvation regimens or undernutrition led to better control of diabetic symptoms. This and later observations served as the basis for the “star-

Table IV.F.1.1. *A historical perspective on dietary recommendations for people with diabetes*

Years	Comments	Carbohydrate	Protein	Fat
Pre-1921	Starvation diets before insulin therapy	Very low	Very low	Very low
1920–1940	Advent of insulin therapy; diets based on food weights, high fat	20%*	10%	70%
1940–1970	First American Diabetes Association (ADA)/American Dietetic Association (ADA) exchange lists (1950)	40%	20%	40%
1971	Revised ADA/ADA dietary recommendations	45%	20%	35%
1986	New ADA/ADA dietary recommendations and exchange lists	≤60%	12–20%	<30%
1994	Individualized diets, less restriction, emphasis on low-fat and high-carbohydrate diet	Variable – but high ≤60%	10–20%	Variable – but <30% with <10% from saturated fat

*Percentage of daily calories.

vation diet.” It is interesting to note that investigators at the Institute for Diabetes, Endocrinology and Metabolic Research in Zagreb recently observed an improvement in glucose control, primarily through weight loss, in populations in the war-torn former Yugoslavia.

Diets in the Age of Insulin Therapy

In 1921, Charles Best and Frederick Banting isolated pancreatic secretions and named them “insulin.” They realized that insulin was responsible for the control of blood glucose levels and the appearance of the clinical symptoms of diabetes. During the 1920s, other endocrinologists characterized the hormones of the pituitary and adrenal glands and indicated that these were also involved in metabolism. However, insulin alone constitutes the antidiabetic hormone therapy in use today for all Type I diabetics and for some Type II diabetics who do not respond to oral hypoglycemic tablets (developed in the 1950s) or diet or exercise. Although it was first used therapeutically in 1922, it was not until the 1950s that appropriate bioassays of insulin were developed.

During the 1920s, Carl Petren advocated a diet high in fat because he was dissatisfied with the starvation and high-protein and high-carbohydrate diets. He deduced that an increase in nitrogen from protein metabolism aggravated acidosis, although we now know that ketoacidosis results from excessive use of fatty acids for energy. He also noted that a high fat intake had a protein-sparing effect that promoted increased carbohydrate use without increases in hyperglycemia and glycosuria. In addition, high-fat diets helped patients maintain their body weight. In the high-fat diet, 70 to 80 percent of the calories were consumed as fat, 10 to 20 percent were consumed as carbohydrates, and 10 percent, or less, as protein.

In the 1930s, H. P. Himsworth exonerated carbohydrates as a cause of worsening metabolic control, which led to a wider acceptance of increased carbohydrate use in diabetic diets. Diets in contemporary times have become increasingly liberal, with a current emphasis on those that are high in carbohydrates and fiber, but contain only moderate amounts of protein, and are low in fats, especially saturated fats (Table IV.F.1.1).

In the 1940s, the American Dietetic Association (ADA), the American Diabetes Association (ADA), and the U.S. Public Health Service sponsored a committee to develop a meal plan or diet that could be easily employed by diabetics. During this time period and well into the 1950s, meal plans were calculated on a daily basis with calories from macronutrients divided into percentages of carbohydrate (40 percent), protein (20 percent), and fat (40 percent).

In the 1950s, the exchange system was developed with its six categories of starches and bread, meat and meat substitutes, vegetables, fruits, milk, and fat.

The exchange lists included foods by weight or volume that contained a specific quantity of calories, carbohydrate, protein, and fat. For example, one exchange from the starch-bread group equaled 80 kilocalories, 15 g of carbohydrate, 3 g of protein, and a trace of fat, whereas one exchange in the meat and meat substitute group for the lean meat category would be the equivalent of 55 kilocalories, 7 g of protein, and 3 g of fat.

Diabetics were taught to adhere rigidly to these exchanges, which were computed on the basis of the total number of calories appropriate for the individual to maintain, lose, or gain weight. Therefore, the number of exchanges in each group would be predicated on the total number of calories prescribed for the patient (for example, a 1,200 or 1,800 calorie diet). Table IV.F.1.1 illustrates the changes in the ADA/ADA recommendations for these proportions. In general, over the years, there has been a decline in the recommended proportion of fats, particularly saturated fats, an increase in carbohydrates, and a moderate decrease in protein. The exchange system did offer flexibility in devising diets that had both variety and palatability. A patient could have, for example, 8 ounces of milk or 8 ounces of yoghurt or 8 ounces of buttermilk, all the equivalent of one milk exchange with approximately the same amount of carbohydrate, protein, and fat content in all three. For a bread-starch exchange, one could have a slice of bread, six saltine crackers, or half a cup of rice, potato, oatmeal, or peas.

Diabetics using insulin therapy are required to eat at regular intervals during the day and to take one or more injections corresponding to food ingestion. The use of both regular and lente, or delayed-action, insulin helps maintain insulin levels to utilize the glucose from food. Various schemes have been proposed for the division of calorie intake throughout the day to correspond to periods of exercise and rest. In general, individuals using insulin therapy often have a snack before retiring. Children are frequently given snacks during late morning and late afternoon, as well as at bedtime. For insulin-dependent diabetics, snacking is especially important before exercise because exercise has an insulinlike effect that can lead to hypoglycemia – a potentially dangerous, even fatal, condition. Nutritionally preferred snacks normally have included a combination of protein, fat, and carbohydrate and explicitly excluded refined sugar. Until very recently, nonnutritive sugar substitutes provided the main sweeteners for “treats” for both children and adults with diabetes.

Dietary recommendations for diabetics in the 1980s, and continuing into the 1990s, have emphasized a decrease in fat intake, primarily because coronary heart disease is more frequent among diabetics, especially those who are obese when compared to age- and sex-matched peers without DM. A few studies that controlled for dietary components have demon-

strated that increases in fat in the diet may be responsible for the movement of people from impaired glucose tolerance to frank diabetes. Recommendations are for diets low in saturated fat, cholesterol, and, more recently, polyunsaturated fats in favor of monounsaturated fats, such as those found in olive oil. There continues to be a renewed “heart-healthy” focus in diabetic diets. Limiting dietary fat decreases the frequently observed hyperlipidemias and enhances weight control in noninsulin-dependent diabetic adults. The World Health Organization (WHO) and various European professional societies recommend a total fat intake equal to or less than 30 percent of total energy intake, saturated fat equal to or less than 10 percent, and the maintenance of serum cholesterol equal to or less than 200 milligrams per deciliter. Because hypertension and vascular diseases, including strokes, are more frequent among diabetics, the recommended salt intake has also been decreased.

Over the last decade, careful attention has been paid to the types of dietary fats, although lowering plasma cholesterol levels has been the primary concern of “heart-healthy” diets. Lipid levels are only partially a reflection of dietary cholesterol and, in fact, are much more responsive to saturated fat intake and to the polyunsaturated-saturated fat ratio. Dairy products (milk, cheese, and yoghurt) contain both cholesterol and saturated fat, which has led to recommendations for lowfat dairy products and fat substitutes in bakery products. Fat substitutes, particularly for weight control, are now common in prepared foods in the United States. Like sugar substitutes, they will most likely become widely distributed and available for home food preparation. Among the different types of dietary fats, stearic acid appears to be a saturated fat that does not raise plasma cholesterol levels. It is common in beef and chocolate.

Polyunsaturated fats play an important role in lipid-lowering diets, particularly the omega-3 fatty acid series in fish oil. Monounsaturated fatty acids, like those in olive oil, have been associated with diets in the Mediterranean region where rates of coronary artery disease are low, yet fat intake is often relatively high. Olive oil is rich in one of the monounsaturated fatty acids, oleic acid, which can lower plasma cholesterol levels. Dietary recommendations may continue to become more specific as we understand the different roles of fats in the body and their contribution to diabetes, coronary heart disease, and obesity. At the moment, the ADA/ADA are recommending 30 percent or less of the total calories from fat and, of this, 10 percent or less from saturated fat, 10 to 15 percent from monounsaturated fat, and the rest from polyunsaturated fatty acids. This may be difficult to achieve for most among the general public, but with proper dietary education for diabetics, such recommendations, it is hoped, will help diminish both the morbidity and mortality associated with heart and vascular diseases.

Diet and Diabetes: Issues in the 1990s

Two nutrition policy and nutrition education initiatives for the U.S. population have had an impact on dietary recommendations for diabetics. These are the Food Guide Pyramid and the new food nutrition labels. For many decades in the twentieth century, the United States focused on the four food groups: bread and starches, meat and meat substitutes, dairy products, and fruits and vegetables. But in 1992, the U.S. Department of Agriculture (USDA) adopted a Food Guide Pyramid, putting foods into five groups plus a high-fat, high-sugar group. In the bread, cereal, rice, and pasta group, individuals are to eat 6 to 11 servings a day. Translated into the exchange system, this is approximately 80 calories per exchange with 15 g of carbohydrate, 3 g of protein, and a trace of fat. From the fruit and vegetable group, individuals are to eat daily 3 to 5 servings of vegetables (at approximately 25 calories a serving with 5 g of carbohydrate and 2 g of protein) and 2 to 4 servings of fruit (at 60 calories a serving with 50 g of carbohydrate). For the milk group, including milk, yoghurt, and cheese, 2 to 3 servings are recommended with an exchange value of 90 calories per serving, 12 g of carbohydrate, 8 g of protein, and 1 g of fat. For the meat, beans, eggs, and nut group, 2 to 3 servings a day are recommended at 55 to 100 calories per serving, 7 g of protein, and 3 to 8 g of fat. At the top of the pyramid, foods from the fats and sweets group are to be used sparingly with approximately 45 calories and 5 grams of fat for each fat exchange. The base of the pyramid is comprised of foods high in carbohydrates and has become the focus of nutrition recommendations for the U.S. population as a whole. The emphasis in the 1990s is to normalize the diet of individuals with diabetes by emphasizing wise choices from the same range of foods recommended to the general population. Thus, the ADA/ADA Exchange lists have been translated into the new Food Guide Pyramid scheme.

Recognizing the need to address ethnic food diversity and in keeping with the goal of individualizing diets, the ADA/ADA has produced a series of ethnic and regional food guides that feature specialty food items, for example, matzo, potato pancakes, and gefilte fish, among traditional and festive Jewish food items, and the tortillas, frijoles, and menudo of Mexican-American diets. Such foods are listed by exchange group and portion size, with a complete accounting of the calories and macronutrients for a portion. In addition, they include amounts of sodium, potassium, and fiber and a division of the fatty acids into saturated, monounsaturated, and polyunsaturated. These food guides are available for both professionals and patients.

After years of debate and deliberation by many interested parties, Americans, beginning in 1993, finally started to see the fruits of that controversy in the form of new food labels on most products carried

by American supermarkets. The new food labels entitled "Nutrition Facts" provide serving size, the number of servings per container, the percent of calories in a serving, and the percent of calories from fat, as well as the total amount of fat and saturated fat, cholesterol, sodium, carbohydrate, dietary fiber, sugar, and protein. These nutrients are also presented as the percent of daily calories for 2,000 and 2,500 kilocalorie diets. In addition, listings on the labels include the percent of the daily requirement for vitamins A and C, along with calcium and iron contained in one serving of the food. The new food labels have been touted as an important point-of-purchase tool for diabetics, enhancing appropriate food choices with a balance of macronutrients, micronutrients, and calories.

Nancy Cooper (1988) summarized the important diet and nutrition goals for individuals with diabetes in the 1990s. These goals were as follows:

1. To restore normal blood glucose and lipid levels, thereby preventing hyperglycemia and/or hypoglycemia and delay the development of long-term cardiovascular, renal, retinal, and neurological complications.
2. To insure normal growth rates for children and adolescents with diabetes and to attain and/or maintain a reasonable body weight for adults.
3. To promote weight loss in obese individuals with changes in food intake and eating behaviors and an increase in physical activity.
4. To develop an individualized meal plan based on diet history, medication protocols, and activity levels that will benefit overall health and control diabetes.

New approaches to reach these goals are based on research that defines the physiological roles of food and the behaviors that involve food consumption. Work beginning in the 1970s by Phyllis Crapo and colleagues and D.J.A. Jenkins and colleagues radically changed ideas about carbohydrates based on the glycemic index of foods. The glycemic index is the area under a blood glucose response curve for a test food, divided by the area under the blood glucose response curve for a standard glucose load $\times 100$. The response curve is developed from blood glucose measurements made both before food ingestion and at 30-minute intervals after, for a duration of 180 minutes. In testing a number of foods, these researchers discovered that the glycemic index for "refined carbohydrates" or sugars was very similar to those of many complex carbohydrates. For example, potatoes and sucrose both have very high glycemic indices, whereas pasta has a much lower glycemic index than either potatoes or bread. Numerous factors affect the glycemic response, such as the fiber and fat in food, cooking and processing of food, and the rate of food ingestion. There are many low glycemic foods such as peanuts, lentils, legumes, pasta, and dairy products. High glycemic foods include refined cereals, breads,

and many root vegetables. Fat in a food will decrease the glycemic index. In the 1980s, researchers began to test meals rather than individual food items and discovered that there was a balance of glycemic indices among foods. Therefore, a food with a high glycemic index can be balanced by foods that have low glycemic indices in meal planning. The other consequence of glycemic index research has been the realization and scientific rationale that it is appropriate to include refined carbohydrates (that is, sugars) in meals because glycemic response studies have shown that equal grams of carbohydrate from sucrose and different types of starches produce similar glucose responses. Sugar, in moderation, is no longer prohibited.

The decade of the 1990s also saw a resurgence of interest in non-Western medical treatments and dietary prescriptions for diabetes. Although American physicians and dietitians have been particularly wary of claims made for dietary supplements, a 1992 issue of the *Journal of the American Diabetes Association*, Florida Affiliate, noted the finding of Richard Anderson that cinnamon increases by ninefold the ability of insulin to metabolize blood glucose. British researchers, particularly S. Swanson-Flatt, P. Flatt, C. Day, and C. Bailey have produced a number of publications evaluating traditional medicines prepared from herbs, spices, and plants. They now have a compendium of more than 700 different species of plants that have been advocated for diabetes treatment.

The efforts of these researchers are particularly noteworthy because they have completed a number of animal experiments, testing individual plants. One notable effect has been the reduction of intestinal glucose absorption because of the high fiber content of many of the plants advocated for diabetes control. Their article in *Proceedings of the Nutrition Society* (1991) is a particularly good summary of their experiments. Some of the herbs and spices they have tested include goat's rue, Java plum, nettle, and sumac. For vegetables they have looked at members of the brassica or cabbage family, lettuce, onion, and potato. They found, for example, that an extract of onion bulb improved glucose tolerance without altering plasma insulin response in normal mice. They also discovered that it exerted a hypoglycemic effect on noninsulin-dependent diabetics. They have tested a number of mushrooms and have found that, in diabetic mice, some of these stimulate insulin release and reduce the severity of hyperphagia, polydipsia, and weight loss.

Many teas made from extracts or infusions of fruits have been tested as well. The fruits have included apple, lime, lemon, raspberry, and blackberry, although most of these have failed to show a significant effect in diabetic animals. Yeast has also been viewed as an important source of chromium and B vitamins that may be deficient among some diabetics. Brewer's yeast has long been held to contain a factor that improves glucose tolerance, but no effect on glycemic control

was evident in their testing of diabetic mice. The British researchers also found that the use of bitter melon or kerela (*Momordica charantia*), which is consumed as a vegetable in Asian diets, improves oral glucose tolerance without altering insulin response. West Indian cerasee and mistletoe both show some glucose-lowering effect. Infusions of alfalfa or lucerne (*Medicago sativa*), used to treat diabetes in South Africa, and eucalyptus leaves have proven somewhat effective in reducing hyperglycemia in mice. The British investigators advocate the use of these hypoglycemic agents as an adjunct to traditional therapy.

Starting with the oat bran craze of the late 1980s in the United States, an increased focus on fiber in the diet of diabetics has also occurred. But prior to that time, J. W. Anderson's research in the 1970s indicated an important hypoglycemic and hypolipidemic effect of fiber in the diet of diabetics. Recent work has focused on elucidating the different types of fiber, particularly digestible and nondigestible fiber, and their impact on glycemic control, cholesterol levels, weight reduction, and gastrointestinal disorders. Contemporary researchers advocate a diet yielding approximately 40 grams of fiber per day, which is about twice the usual intake of the U.S. population. Although the efficacy of such a regimen has been shown in clinical studies, it may be difficult for individuals to modify their food intake sufficiently to obtain this amount of fiber and still eat a palatable meal.

Most research has not found a causal link between the onset of diabetes and diet (for example, sugar intake). However, several recent studies have indicated a link between cow's milk ingestion, the production of antibodies against casein (cow's milk protein), and the onset of Type I diabetes. Investigators suggest that this immune response can trigger an autoimmune response against the insulin-producing cells of the pancreas, destroying them and causing insulin-dependent diabetes mellitus. Although other researchers have not found such a link, cow's milk allergy is common among children.

Research on the role of breast feeding in the etiology of Type I diabetes has produced evidence both confirming and refuting the relationship. The arguments focus on the immunity conferred upon the infant by mother's milk, as well as the protection from immunological responses that could be catalyzed by cow's milk or cow's milk-based infant-feeding formulas, especially in very young infants.

Finally, we should note that Gerald Reaven (1992) has proposed that prenatal and early postnatal undernutrition, followed by overnutrition later in life, can also pose a risk for the onset of diabetes. He suggests that this disruption of normal embryogenesis produces an underlying mechanism of insulin insufficiency and/or glucose intolerance that can have a contributory effect on micro- and macrovascular disease and hypertension, as well as diabetes. He calls this Syndrome X.

Obesity has been a focus for the 1990s. Approximately 70 to 90 percent of adults are obese at the time diabetes is diagnosed. Weight reduction and/or weight control have been a primary goal of nutritional intervention. Unfortunately, the success rates of such nonpharmacological therapies have been very low. Most of the interventions have focused on behavior modification to reduce caloric intake and increase caloric expenditure. Special prescription diets are recommended for individuals who are showing the sequela or complications of diabetes, particularly low-sodium diets for those with hypertension and lower-protein diets for those with renal disease. The fourth population for which there are extensive dietary prescriptions are Type I diabetics and gestational diabetics who need special dietary control during pregnancy to insure healthy outcomes for their infants and themselves. During the 1990s, a greater emphasis has been placed on prophylactic diets to prevent obesity and the complications of diabetes.

In summary, the 1994 ADA/ADA recommendations set dietary goals for people with diabetes that are, generally, the same as those for the general population. The aim is to maintain a normal weight, reduce fat, increase fiber, reduce salt, and limit alcohol intake. The use of nonnutritive sweeteners and fats has been recommended. The new emphasis on the liberalization of diets, the availability of new foods, nonnutritive sweeteners, and fats will help to change the therapeutic diet from one that is only moderately palatable to one that is both desirable and efficacious in achieving the dietary goals for people with diabetes.

Leslie Sue Lieberman

The author wishes to thank Sarah Gross, M.S., R.D., and G-Hyon Gang, M.A., Ph.D., for their contributions to this chapter.

Bibliography

- American Diabetes Association. 1987. Nutritional recommendations and principles for individuals with diabetes mellitus. *Diabetes Care* 10: 126-32.
- 1991a. Position statement: Nutritional recommendations and principles for individuals with diabetes mellitus. *Diabetes Care* 14: 20-7.
- 1991b. Position statement: Use of noncaloric sweeteners. *Diabetes Care* 14: 28-9.
- 1994. Position statement: Nutritional recommendations and principles for people with diabetes mellitus. *Diabetes Care* 17: 519-22.
- American Diabetes Association and American Dietetic Association. 1987. *Exchange lists for meal planning*. Alexandria, Va.
- Anderson, J. W., and A. O. Akanji. 1991. Dietary fiber - an overview. *Diabetes Care* 14: 1126-31.
- Anderson, J. W., and K. Ward. 1979. High carbohydrate, high fiber diets for insulin treated men with diabetes. *American Journal of Clinical Nutrition* 21: 2312-21.
- Arætaeus. 1856. *The extant works of Arætaeus, the Cappadocian*, ed. and trans. Francis Adams. London.

- Baker, D. J. P. 1994. *Mothers, babies, and disease in later life*. London.
- Bennion, M. 1979. *Clinical nutrition*. New York.
- Bernard, Claude. 1855. Sur le mécanisme de la formation du sucre dans le foie. *Centre de la Recherche de l'Académie de Science* 41: 46-9.
- Bliss, Michael. 1982. *The discovery of insulin*. Chicago.
- Boyd, Eaton. 1988. *The paleolithic prescription: A program of diet and exercise and a design for living*. New York.
- British Diabetic Association. n.d. Dietary recommendations for the 1980s (manuscript). London.
1992. Dietary recommendations for people with diabetes: An update for the 1990s. *Diabetic Medicine* 9: 189-202.
- Campbell, Donald. 1973. *Arabian medicine and its influence on the middle ages*. New York.
- Chevreul, M. E. Eugène. 1815. Note sur le sucre de diabète. *Annuaire de la Chimie* 95: 319-20.
- Committee on Diet and Health, Food and Nutrition Board, Commission on Life Sciences, National Research Council. 1989. *Diet and health: Implications for reducing chronic disease risk*. Washington, D.C.
- Cooper, Nancy A. 1988. Nutrition and diabetes: A review of current recommendations. *The Diabetes Educator* 14: 428-32.
- Dash, Bhagwan. 1989. *Ayurvedic cures for common diseases*. New Delhi.
- Dash, Bhagwan, and Manfred M. Junius. 1983. *A hand book of Ayurveda*. New Delhi.
- Day, C. 1990. Hypoglycaemic compounds from plants. In *New antidiabetic drugs*, ed. C. J. Bailey and P. R. Flatt. London.
- Diabetes Care and Education Dietetic Practice Group, American Diabetic Association. 1989a. *Jewish food practices, customs, and holidays. Ethnic and regional food practices: A series*. Alexandria, Va.
- 1989b. *Mexican American food practices, customs, and holidays. Ethnic and regional food practices: A series*. Alexandria, Va.
- Diabetes Control and Complications Trial (DCCT) Research Group, The. 1993. Nutrition interventions for intensive therapy in the diabetes control and complications trial. *Journal of the American Dietetic Association* 93: 765-72.
- Dobson, Matthew. 1776. Experiments and observations on the urine in diabetes. *Medical Observations and Inquiries* 5: 298-316.
- Ebstein, E. 1989. Diabetes, sugar consumption and luxury throughout the ages. In *Diabetes - its medical and cultural history*, ed. D. Von Engelhardt, 101-6. New York.
- Estes, J. Worth. 1989. *The medical skills of ancient Egypt*. Canton, Mass.
- Franz, J. J., E. S. Horton, J. P. Bantle, et al. 1994. Nutrition principles for the management of diabetes and related complications. *Diabetes Care* 17: 490-518.
- Gerstein, H. C. 1994. Cow's milk exposure and type I diabetes mellitus. A critical review of the clinical literature. *Diabetes Care* 17: 13-19.
- Grandjean, Ann C., and Carol J. Bennett. 1980. *Eat less food*. Omaha, Neb.
- Grundy, Scott M. 1991. Dietary therapy in diabetes mellitus. Is there a single best diet? *Diabetes Care* 14: 796-801.
- Hazlett, Lucille. 1992. From raw quinces to gruel to insulin: A brief history of diabetes discovery and treatment. *Diabetes Interview* 2: 9.
- Himsworth, H. P. 1935. The dietetic factor determining the glucose tolerance and sensitivity to insulin of healthy men. *Clinical Science* 2: 67-94.
- 1935-6. Diet and the incidence of diabetes mellitus. *Clinical Science and Molecular Medicine* 2: 117-48.
- Hippocrates. 1949. *Ancient medicine and other treatises*, ed. and trans. Francis Adams. Chicago.
- Hollenbeck, Claire B., and Ann M. Coulston. 1991. Effects of dietary carbohydrate and fat intake on glucose and lipoprotein metabolism in individuals with diabetes mellitus. *Diabetes Care* 14: 774-85.
- Huh, Joon. 1596. TongUiPoKam [collection of classified medical prescriptions], translated into current Korean language in 1966 by YoungHoo Kim, Kilgu Shin, Jaesung Kim, and Wonshic Bae. Seoul.
- Jenkins, D. J. A. 1982. Lente carbohydrate: A new approach to dietary management of diabetes. *Diabetic Care* 5: 634-41.
- Jenkins, D. J. A., T. M. S. Wolever, S. Bacon, et al. 1980. Diabetic diets: High carbohydrate combined with high fiber. *American Journal of Clinical Nutrition* 33: 1729-33.
- Joslin, Elliott Proctor. 1916. *Treatment of diabetes mellitus*. Philadelphia, Pa.
1921. The prevention of diabetes mellitus. *Journal of the American Medical Association* 76: 79-84.
1994. *Joslin's diabetes mellitus*. Thirteenth edition, ed. G. Ronald Kahn and Gordon C. Weir. Philadelphia, Pa.
- Journal of the American Diabetes Association/Florida Affiliate*. 1992. Cinnamon, a health benefit. *Diabetes Journal* 2: 3.
- Krause, M. V., and L. K. Mahan. 1979. *Food, nutrition and diet therapy*. Sixth edition. Philadelphia, Pa.
- Leeds, A. R. 1979. The dietary management of diabetes in adults. *Proceedings of the Nutrition Society* 38: 365-71.
- Lepine, R. 1909. *Le diabète sucre*, ed. F. Alcan. Paris.
- Levine, R. 1986. Historical view of the classification of diabetes. *Clinical Chemistry* 32 (Suppl.): 84-6.
- Lieberman, Leslie Sue. 1993. Diabetes. In *The Cambridge world history of human disease*, ed. K. Kiple, 665-76. New York and Cambridge.
- Lu, Gwei-Djen, and Joseph Needham. 1980. *Celestial lancets: A history and rationale of acupuncture and moxa*. Cambridge.
- Marble, A. 1989. John Rollo. In *Diabetes - its medical and cultural history*, ed. D. Von Engelhardt, 230-4. New York.
- Marshall, J. A., S. Hoag, S. Hetterly, et al. 1994. Dietary fat predicts conversion from impaired glucose tolerance to NIDDM: The San Luis Valley diabetes study. *Diabetes Care* 17: 50-6.
- Maryniuk, Melinda D. 1989. Hyperlipidemia and diabetes: The role of dietary fats. *The Diabetes Educator* 15: 258-64.
- McQuillan, J. E., and M. S. McQuillan. 1982. Discovery of insulin and control of diabetes mellitus. *Janus* 69: 97-118.
- Medvei, Victor C. 1982. *A history of endocrinology*. Lancaster, England.
- Nathan, David M. 1987. The glycemic index: Meat and potatoes or just gravy. *Diabetes Care* 10: 524-5.
- National Diabetes Data Group. 1979. Classification and diagnosis of diabetes mellitus and other categories of glucose intolerance. *Diabetes* 28: 1039-57.
- National Institutes of Health. 1987. Consensus development conference on diet and exercise in non-insulin-dependent diabetes mellitus. *Diabetes Care* 10: 639-44.
- Nuttall, F. 1993. Dietary fiber in the management of diabetes. *Diabetes* 42: 503-8.

- Pan, David A., A. J. Hulbert, and L. H. Storlien. 1994. Dietary fats, membrane phospholipids, and obesity. *Journal of Nutrition* 124: 1555–65.
- Powers, M. A. 1988. *Nutrition guide for professionals – diabetes education and meal planning*. Alexandria, Va.
- Rabinowitch, J. M. 1935. Effects of the high carbohydrate – low calorie diet upon carbohydrate tolerance in diabetes mellitus. *Canadian Medical Association Journal* 26: 136–44.
- Rapp, Stephen, Patricia Dubbert, and Paul Burkett. 1986. Food portion size estimation by men with Type II diabetes. *Journal of the American Dietetic Association* 86: 249–51.
- Reaven, Gerald M. 1992. Syndrome X. *Blood Pressure Supplement* 4: 13–16.
- Saundby, R. 1908. Diabetes mellitus. In *A system of medicine by many writers*, ed. C. Allbutt and H. D. Rolleston, 167–212. London.
- Schadewaldt, H. 1989. History of diabetes mellitus. In *Diabetes – its medical and cultural history*, ed. D. Von Engelhardt, 405–7. New York.
- Schumacher, H., and J. Schumacher. 1989. Then and now: 100 years of diabetes mellitus. In *Diabetes – its medical and cultural history*, ed. D. Von Engelhardt, 246–7. New York.
- Shin, Jae Yong. 1989. *DangRyoByong gua sogal* (Diabetes, and the disease of exhaustion and thirst). Seoul.
- Shuman, C. R., and I. L. Spratt. 1981. Office guide to diagnosis and classification of diabetes mellitus and other categories of glucose intolerance. *Diabetes Care* 4: 335.
- Striker, Cecil. 1961. *Famous faces in diabetes*. Boston.
- Swanston-Flatt, S., P. R. Flatt, C. Day, and C. J. Bailey. 1991. Traditional dietary adjuncts for the treatment of diabetes mellitus. *Proceedings of the Nutrition Society* 50: 641–51.
- Taylor, R., and P. Zimmet. 1983. Migrant studies and diabetes epidemiology. In *Diabetes in epidemiological perspective*, ed. J. I. Mann, K. Pyorala, and A. Teuscher, 58–77. Edinburgh.
- The Papyrus Ebers. 1937. *The Papyrus Ebers: The greatest Egyptian medical document*. Copenhagen.
- Tinker, Lesley Fels, Joan M. Heins, and Harold S. Holler. 1994. Commentary and translation: 1994 nutrition recommendations for diabetes. *Journal of the American Dietetic Association* 94: 507–11.
- Trowell, H., D. P. Burkitt, and K. Heaton. 1985. *Dietary fiber, fiber-depleted foods and disease*. London.
- U.S. Department of Agriculture. 1990 (rev. 1995). *Dietary guidelines for Americans*. Hyattsville, Md.
- Von Engelhardt, D. 1989. Outlines of historical development. In *Diabetes – its medical and cultural history*, ed. D. Von Engelhardt, 3–10. New York.
- Von Mering, J., and O. Minkowski. 1889. Diabetes mellitus nach Pankreas Stirpation. *Zentralblatt für klinische Medizin* 10: 393–4.
- Welch, Robert W. 1991. Diet components in the management of diabetes. *Proceedings of the Nutrition Society* 50: 631–9.
- West, K. M. 1978. *Epidemiology of diabetes and its vascular lesions*. New York.
- Wilder, R. M. 1989. Karl Petren. A leader in pre-insulin dietary therapy of diabetes. In *Diabetes – its medical and cultural history*, ed. D. Von Engelhardt, 405–7. New York.
- Willis, Thomas. 1684. On excessive pissing and the remedy for it particularly of the diabetes (or pissing disease) into the theory and cure for which we here inquire. In *Pharmaceutics rationalis*, 71–6. London.

Wrenshall, Gerald Alfred. 1962. *The story of insulin: Forty years of success against diabetes*. London.

Wylie-Rosett, J., and H. Rifkin. 1985. The history of nutrition and diabetes. In *Nutrition and diabetes*, ed. L. Jovanovic and C. M. Peterson, 1–13. New York.

Zeman, F. J. 1983. *Clinical Nutrition and Dietetics*. Lexington, Mass.

IV.E2 🐉 Nutrition and Cancer

As early as the 1930s, experiments on laboratory animals revealed that diet can considerably influence the process of cancer causation and development (carcinogenesis) (Tannenbaum 1942a, 1942b; Tannenbaum and Silverstone 1953). It was several decades later, however, that the first epidemiological studies appeared to indicate that diet could play a role in human cancer. A key conference held in 1975, entitled “Nutrition in the Causation of Cancer,” summarized the existing knowledge and hypotheses (Wynnder, Peters, and Vivona 1975). From that moment, research in experimental systems, including animal models and epidemiological studies, increased rapidly, providing extensive information on the impact of nutritional traditions and specific macro- and micronutrients on several types of cancer. Considerable progress had already been made in several underlying sciences. For example, advances had been achieved in understanding the mechanisms of action of nutrients, the process of carcinogenesis, and the classification of carcinogens according to their mode of action (Kroes 1979; Weisburger and Williams 1991).

In particular, epidemiological studies on the international variations in incidence rates for certain cancers pointed to the existence of one or more exogenous factors that could be controlled. Observational studies had been conducted with migrants from countries with lower incidence rates to countries with higher incidence rates. A rapid increase from the lower to the higher incidence in those migrants supported the suggestion that environmental causes, and especially prevailing dietary habits, may influence the development of a number of neoplasms.

For example, in the Western world, high incidences of cancers in the lung, colon and rectum, breast, prostate, pancreas, endometrium, and ovary were observed. But these diseases were, until recently, quite rare in Japan and other Asian countries, where cancers of the stomach, esophagus, and liver were the major diseases. Yet, in migrant populations the risk of cancer shifted from the risk prevalent in their native country to that seen in the new host country, sometimes within the same generation (colon), sometimes in the next generation (breast) (Parkin 1993).

The results of experimental and epidemiological studies have led to hypotheses concerning factors

involved in cancer causation. In fact, E. L. Wynder and G. B. Gori (1977) and R. Doll and R. Peto (1981) estimated that 35 percent of cancer mortality may be attributable to diet, thereby placing diet at the same "risk factor" level as smoking, which was calculated to cause about 30 percent of cancer mortality. In fact, current views suggest that dietary habits may account for 55 to 60 percent of cancer mortality. The interpretation of such multidisciplinary investigations has shed light on underlying mechanisms of cancer causation and has provided an understanding of the role of specific dietary risk factors.

In this chapter, the focus is the role of diet in cancer causation. First, attention is drawn to aspects of mechanisms in carcinogenesis. The major strengths and weaknesses of experiments in animal models and epidemiological studies are discussed, and then the major diet-related causes are examined and their suggested mechanisms and possible preventive measures described. The safest (but in the public's perception the most dangerous) intentional additives in food are evaluated for their cancer risk, as are nonintentional contaminants, along with substances that occur naturally in food. Finally, the key role that diet may play in an integrated approach to (chronic) disease prevention is presented. Health promotion through research-based nutrition not only is important for the individual but, on a global level, may also appreciably lower the cost of medical care. A research-based healthy lifestyle is designed to allow people to "die young, as late in life as possible," as Dr. Ernst L. Wynder, president of the American Health Foundation, has said.

Tumor Models: Strengths and Weaknesses

Animal tumor models are often used in experimental designs to study the mechanisms of cancer causation, to examine the effects of modulating factors on the genesis and development of cancer, to assess therapeutic modalities, and to explore possible adverse effects.

Tumor models are specifically used to investigate etiologic and physiopathological properties or processes, especially those which, for obvious practical and ethical reasons, cannot be studied in humans. The ideal animal tumor is histologically similar to the human neoplasm of concern, and latency period, growth, and tendency to metastasize should both be predictable and resemble those of the human neoplasm. The animals should be cost effective, be easily available, and have a genetic uniformity. Variables should be controllable, thereby making it possible to investigate the influence of isolated factors (Davidson, Davis, and Lindsey 1987; Galloway 1989; Weisburger and Kroes 1994).

For research into matters of nutrition and cancer, chemically induced tumors in animals, predominantly those in rats and mice, are the most important models because they usually best mimic existing types of

human cancer (Kroes et al. 1986). For most, if not all, nutrition-related human cancers, tumor model systems are available. They provide an ideal research tool for investigating the influence of individual factors, either in the initiation phase or in the promotion-progression phase of tumor development, or in the overall process.

Possible leads from epidemiological studies can be further investigated in models, thus providing more detailed information, especially regarding risk factors and mechanisms that can be the basis for new epidemiological studies to test a presumed hypothesis. However, animal tumors are only approximations that are rarely identical to human disease. For example, relatively high dosages of genotoxic carcinogens are used to induce the tumor, which tends to distort circumstances when modulating factors are investigated.

Tumor metabolism, growth, potential for and pattern of metastasis, and clinical features can also differ from the human disease. Multiple tumors often occur in models, but seldom in humans. Moreover, the nutrition patterns of experimental animals can be quite distinct from those of humans. Nonetheless, diets can be designed in terms of macro- and micronutrient intake that mimic specific human nutritional traditions.

Animal studies are best understood as providing support for epidemiological studies. It is the integration of results from epidemiological and animal studies that provides the best insight into the etiology and growth of cancer, as well as its treatment, and brings us nearer to the ultimate goal of such research, cancer prevention.

Epidemiological Studies Concerning Nutrition and Cancer

Epidemiological research on cancer and diet seeks to associate exposure to certain dietary factors with the occurrence of cancer in selected population groups. Studies can be descriptive, reporting the occurrence of cancers in populations, in subgroups of a given population, or in a certain population over time. Observed patterns may be related to particular variables, such as diet, but although the results of such investigations are suggestive, they are certainly not definitive. Their use is especially valuable, however, in identifying populations at risk.

Correlation studies investigate the possible relationships between more-or-less crude exposure data and cancer incidence data in different populations or nations in order to generate new hypotheses. But they are generally of limited value, because national per-capita food intake and cancer incidence data are only approximations and differ from place to place in terminological definition and accuracy. Results should be used mainly as an indication of trends or relationships.

Another type of investigation is the case control

study, in which the investigator is able to collect data from individuals instead of groups, and confounding variables can be controlled to a certain extent. In studies of a specific type of cancer, food intake data are collected and are compared to data similarly obtained in matched controls. When known biases and chance can be excluded, associations can be made between exposure and disease. Case control studies are relatively cheap and are of short duration but never prove causal relationships.

Much more expensive, time consuming, and elaborate are prospective cohort studies that focus on individuals in a given (large) population and establish exposure or nutrition data before the occurrence of the disease. Thus, at the time of disease occurrence, exposure data of the patients are compared to data from people not having the disease, thereby providing evidence for a possible causal relationship. Drawbacks in such studies are that the exposure assessment at any given time may not be representative of the individual's whole life, and this is especially likely in the case of diet.

Finally, intervention studies provide the investigator with the possibility of a random assignment of subjects of a given (often high-risk) population into groups that are treated or fed differently under controlled conditions. In this type of trial, causal relationships can be established.

In investigations focused on the relationship between nutritional factors and cancer, the methods used for determining dietary intake are crucial and difficult. All dietary intake measures share certain limitations, because people vary in their abilities to estimate the amount of something they have eaten. Indeed, sometimes they may even fail to notice or report consumption of certain foods, and they usually possess insufficient knowledge about the ingredients in the foods they consume. A further handicap is that accuracy in dietary recall deteriorates over time. Finally, in the case of cancer, it is very difficult to relate varying diets to the disease because the latter has a long course of development before becoming clinically manifest (10 to 40 years).

The major strength of epidemiological studies, however, is their focus on human populations, thus avoiding the need to extrapolate from other species. They tend to afford an opportunity to examine different effects at different exposure levels, and they are always realistic, in contrast to the high exposure levels usually employed with animals (National Research Council 1982; International Agency for Research on Cancer 1990; Weisburger and Kroes 1994). In this connection, however, it should be noted that in specific investigations into the role of nutrients such as fat, the experimental design in laboratory animals usually faithfully mimics the situation of human populations at high or low risk, thus providing relatively reliable comparative data. But at the same time, as noted, it should also be recognized that the normal food pat-

terns of laboratory animals differ from those of humans.

Mechanisms of Carcinogenesis

The concept that chemicals can induce cancer through a variety of modes of action is derived from a greater understanding of the complex processes of carcinogenesis (Williams and Weisburger 1991). Cancer causation and development involves a series of essential steps (Figure IV.F.7.1). In the first step, a reactive form of carcinogen (often produced metabolically from a procarcinogen) binds to DNA, or DNA is altered by the effective generation of hydroxy radicals. This reaction, in turn, leads to translocation and amplification of specific genes, proto-oncogenes, or a mutation in tumor suppressor genes, that translate to a distinct expression of the properties of the cells bearing such altered genes (Williams and Weisburger 1991; Ronai 1992; Miller 1994).

The property to bind to DNA is the basis for the development of specific rapid, efficient, and economical *in vitro* bioassays, such as a mutation assay in prokaryotic or eukaryotic cell systems. Advantage can also be taken of the presence of enzyme systems performing DNA repair; these can provide effective complementary test systems to outline the possible DNA-reactivity of chemicals (Weisburger and Williams 1991; Weisburger 1994). Moreover, the ³²P-postlabeling procedure of K. Randerath and colleagues (1989) yields information about the presence of reactive carcinogen-DNA adducts.

When a chemical displays properties of reacting with DNA and inducing mutations and DNA repair in a number of cell systems, it can be considered DNA-reactive or genotoxic. Most human carcinogens are genotoxic. Other agents, such as the hormonoid diethylstilbestrol (DES) or the hormone estradiol, may give rise to genotoxic products (Liehr 1990). The fact that a given product is genotoxic may signal that, with sufficient dosage and chronicity of exposure, it poses a cancer risk to humans.

Reactive oxygen, hydroxy radicals, or hydrogen peroxide can be generated in metabolic processes. On the other hand, there are endogenous defense mechanisms, such as catalase-destroying hydrogen peroxide or superoxide dismutase or glutathione peroxidase (in turn, involving glutathione [GSH] or sulfhydryl amino acids), that neutralize reactive oxygen. Exogenous antioxidants (such as those in vegetables, tomatoes, and tea), carotene, vitamins C and E, selenium, and sources of GSH aid in the disposition of reactive species (Harris 1991; Olson and Kobayashi 1992; Packer 1992; Sai et al. 1992).

The consequences of DNA-carcinogen interactions are beginning to be explored in terms of proto-oncogene – oncogene codon translation and amplification that can be measured by highly specific and sensitive polymerase chain reaction techniques

(Brugge et al. 1991). Errors may be introduced into DNA during biosynthesis (Echols and Goodman 1991). Mutational events in tumor-suppressor genes also yield abnormal DNA, which is of growing interest (Brugge et al. 1991). The rate of cell duplication is important in generating abnormal DNA. A rapid rate decreases the chances of successful repair and is a reason that growing organisms or proliferating tissues are often more sensitive to carcinogens (Cohen and Ellwein 1992). For example, radiation exposure during the Hiroshima atom-bomb explosion caused a fourfold higher breast-cancer incidence in the 10- to 15-year-old age group (with mammary gland cells in rapid DNA synthesis and mitosis) than in younger or older groups (Land et al. 1980).

Reactive carcinogens modify not only DNA but also proteins. Both types of interactions can serve as sensitive markers for qualitative and quantitative analysis, especially the readily measured hemoglobin adducts (Brugge et al. 1991).

Nongenotoxic carcinogens and promoters cannot cause cancer without an antecedent mutational event and cell change. For example, mice with mammary tumor virus (MTV) develop mammary tumors proportional to the level of estrogen administered, but those without MTV do not, no matter what dose of estrogen is used (Highman, Norvell, and Shellenberg 1977). The action of promoters requires their presence at relatively high levels for a long time, and that action is often tissue specific. For example, bile acids are promoters of colon cancer, and very high dosages of sodium saccharin act as a promoter for cancer of the urinary bladder. The interruption of gap junctions and intercellular communication plays a key role in promotion, and promoters can be detected through this characteristic (Yamasaki 1990; Trosko and Goodman 1994).

As promoters are nongenotoxic substances, linear extrapolation for health-risk assessment seems unrealistic and, actually, scientifically improper (Kroes 1987; Williams and Weisburger 1991; Weisburger 1994). In order to better assess risk using appropriate epidemiological and biostatistical approaches, new procedures to define the mode of action of epigenetic (nongenotoxic) agents are being developed. Most likely, dose-response studies will yield a typical pharmacological S-shape response, with a definite no-effect level. This is especially important when fundamental insights into the properties of carcinogens and promoters are applied to the area of nutritional mechanisms in cancer causation.

Thus, chemical carcinogens can be classified into two main groups: (1) DNA-reactive substances that are genotoxic in appropriate test systems and (2) epigenetic (nongenotoxic) agents operating by producing some other specific biological effect as the basis for their carcinogenicity. Genotoxic carcinogens alter DNA, are mutagenic, and lead to transformed cells with neoplastic attributes. Nongenotoxic carcinogens involve other mechanisms, such as cytotoxicity,

chronic tissue injury, hormonal imbalances, immunologic effects, or promotional activity.

Diet-Related Cancers

Several extensive reviews addressing diet-related neoplasms have been published in the past decades. In the Western world, cancers associated with nutrition account for a substantial percentage – about 35 to 45 percent – of premature deaths. In this section, characteristics of the different types of cancer, their established or suggested relationships with dietary factors, and their presumed mechanisms of action are described. In addition, possible measures to prevent or decrease the risk of developing the disease are discussed.¹

Oral Cavity, Pharynx, and Esophagus

Cancers of the oral cavity and pharynx account for approximately 400,000 new cancer cases each year in the world. High incidences are noted in France, Switzerland, northern Italy, Central and South America, parts of Pakistan, and India. These cancers occur much more frequently in males than females, and differences between high- and low-incidence areas may be as much as 20-fold.

High-incidence rates of cancer of the esophagus are found in the so-called Asian esophageal cancer belt, which extends from eastern Iran, along the Caspian Sea, through Turkmenistan, Tajikistan, Uzbekistan, and Kyrgyzstan, and into parts of China. Except for the high-incidence areas, where the sex ratio almost equals 1:1, males show predominantly higher incidences. World incidence rates differ more than 100-fold, and globally more than 300,000 new cases occur each year.

Cancer of the esophagus is especially common among individuals who chew or smoke tobacco and drink alcoholic beverages. Consumption of alcohol alone – especially hard liquor – seems to be a risk factor as well (Seitz and Simanowski 1991; Castelletto et al. 1994). Smoking and alcohol consumption have a synergistic effect on carcinogenesis in the upper alimentary tract. In Asian and African populations, dietary deficiencies of zinc, riboflavin, vitamins A and C, manganese, and molybdenum may play a role, as well as mycotoxins, bracken fern, opium pyrolysates, and betel quids. The consumption of salted fish is an established risk factor in southern Chinese populations, probably because of the formation of specific nitrosamines (Craddock 1992; Zeng et al. 1993). Consumption of very hot beverages, along with the use of substances that irritate the oral cavity, pharynx, and esophagus, all of which lead to increased cell proliferation, may enhance the incidence of neoplasia.

However, substantial differences in incidence between high- and low-risk areas indicate that there exists considerable potential for prevention. Frequent consumption of fresh fruits and vegetables, as well as

tea, appears to be associated with a lower risk for these types of cancer. The potential reduction has been estimated to be around 75 percent (Negri et al. 1993). Preventive measures involve the avoidance of tobacco and very hot beverages, along with moderate alcohol use and a well-balanced diet that includes a sizable increase in the regular consumption of vegetables and fruits (Block, Patterson, and Subar 1992).

Stomach

In the 1980s, gastric cancer was still considered to be the most common cancer in the world. Indeed, with almost 700,000 new cases per year, it represented approximately 10 percent of all cancers. Differences between high- and low-incidence areas vary by 40-fold. However, a large decrease in rates has occurred in most populations during the last four to five decades, indicating a reduction in exposure to tissue-specific carcinogens and/or the introduction of a protective agent.

Males suffer approximately twice the incidence and mortality of females, although the sex ratio is not constant by age group. The sex ratio equals 1:1 in people under the age of 30, but the disease is rare in this group. High-risk populations usually consume considerable quantities of pickled vegetables, dried salted fish, smoked fish, and other smoked, salted, and dried foods. Consumption of certain salted and pickled fish has yielded high levels of mutagenicity and evidence of carcinogenicity. One of the mutagens present has been identified as 2-chloro-4-methylthiobutanoic acid. This finding was totally unexpected, because in the past, nitroso compounds were associated with stomach cancer (Chen et al. 1995).

By contrast, a negative association has been established between mutagenicity and the regular intake of green leafy vegetables and citrus fruits. Laboratory experiments show that vitamins C and E block the formation of mutagens when fish is treated with nitrite, mimicking pickling (Weisburger 1991).

Infection with *Helicobacter pylori* and associated conditions, such as atrophic gastritis, ulceration, partial gastrectomy, bile acid reflux, and pernicious anemia, are additional risk factors. Several of these increase cell-duplication rates, rendering the gastric cells more sensitive to genotoxic carcinogens.

A high level of consumption of salted, pickled, or smoked food was once customary in the Western world. However, better access to home refrigeration, improved and cheaper transport – and, therefore, increased availability – of fresh fruit and vegetables seems to correlate well with the decline of this type of cancer (Howson, Hiyama, and Wynder 1986; Weisburger 1991). The relevant mechanism begins with the development of atrophic gastritis due to the cytotoxic activity of salt and vitamin deficiencies. The consequent decrease in gastric acidity permits uninhibited bacterial growth. Bacterial growth then converts dietary nitrates to nitrites, which are further metabolized into carcino-

genic nitroso derivatives or reactive carcinogens (Correa 1992; Chen et al. 1995). Because vitamins C and E are known to be effective inhibitors of nitrosation, it is plausible that an increased intake of these vitamins, or foods containing them, should reduce the risk of gastric cancer by inhibiting nitrosation.

Preventive measures are the introduction of food refrigeration, the reduction of salt and pickled food intake, and an increased consumption of fruits and vegetables. Especially in areas with high prevailing environmental nitrate levels, vitamin C and vitamin E supplementation may be useful for preventing formation of nitrite-derived reactive carcinogens and reducing nitrite produced by conversion of nitrate in the mouth. The preventive potential has been estimated to be about 50 percent but may well be much higher.

Colon and Rectum

Approximately 600,000 new cases of colorectal cancer are diagnosed worldwide each year. It is particularly a disease of the developed countries, which to some extent reflects increasing life expectancy. Differences in incidence may be 60-fold. The lowest incidence rates are found in Africa and Asia, although incidences are rising, especially in areas where the risk was formerly low, as in Japan. Colon cancer affects the sexes equally. The distribution for rectal cancer is similar to that for colon cancer. Incidences are usually lower, and there is a male-female ratio of 1.5:2.0, especially in high-incidence areas.

Epidemiological evidence suggests that (Western) lifestyle is an important determinant of risk for colorectal cancer: Migrants to Western countries acquire a higher risk for the disease in the first generation, and Mormons and Seventh Day Adventists enjoy a low risk. Familial polyposis, ulcerative colitis, and Crohn's disease are identified risk factors for colon cancer, but these are uncommon conditions.

Diets high in fats and low in fiber and vegetables are associated with increased risk for colon cancer. A fat-fiber interaction has been suggested, and the type of fiber is important as well (Kroes, Beems, and Bosland 1986; Weisburger 1992). It is interesting to note that some polyunsaturated fats found in fish and some vegetable seeds inhibit colon cancer formation. Moreover, olive oil intake, as in the Mediterranean countries, does not increase the risk of the nutritionally linked cancers or heart disease, a fact also documented in animal models (Reddy 1992). An inverse relationship has also been found for the consumption of fruits and vegetables, as well as for calcium intake and regular exercise; the same is true of coffee and tea for colon and rectal cancer, respectively (Baron, Gerhardson de Verdier, and Ekblom 1994).

Experimental and epidemiological research has revealed that bile acids promote cancer formation (Reddy 1992). The case is similar with alcohol, especially for rectal cancer, perhaps accounting for the higher male-to-female ratio (Seitz and Simanowski

1991). Recent surveys also indicate that intake of heavily fried or grilled meat and gravies is positively related to colorectal cancer, suggesting that chemicals produced during the frying or grilling of meats (heterocyclic aromatic amines) may be the initiating carcinogens, particularly for breast, colon, and, perhaps, prostate and pancreatic cancer (Adamson et al. 1995).

Suggested mechanisms in colon cancer development are the increased bile acid concentrations in individuals consuming high levels of many types of dietary fat. The higher concentrations of bile acids may lead to increased turnover of the epithelial cells of the intestines, reflecting increased risk of carcinogen-DNA adducts to cause translocation and amplification of abnormal genes or mutated tumor-suppressor genes. This phenomenon is inhibited by increased dietary calcium. The toxicity of bile acids is also reduced at a lower luminal pH. Alcohol may act as a promoter at the level of the rectum, and its metabolite acetaldehyde, which occurs at higher concentration in the rectum, may induce cytotoxicity, thus leading to increased cell proliferation and turnover.

Dietary fibers at adequate concentrations are thought to dilute and particularly bind the genotoxic agents present in the gut, decreasing fecal mutagenic activity. Fibers also modify the metabolic activity of the gut flora and lower luminal pH. Wheat bran increases the bulk of the gut contents, thus diluting bile acids and decreasing their adverse effect on the mucosal lining of the bowel (Reddy 1992). Regular physical exercise also lowers transit time of the luminal contents and appears to decrease risk of colorectal cancer. Fruits and vegetables generally reduce cancer risk through several mechanisms (Block et al. 1992). They provide fibers and antioxidants that can detoxify active genotoxins and also contain a number of chemopreventive agents, such as indole derivatives, that are anticarcinogenic. Tea antagonizes the effect of heterocyclic amines present in fried or broiled meats, which are thought to be carcinogens for the colon.

Potential reduction of colorectal cancers through prevention has been estimated at 35 percent, mainly of distal colon cancer and rectal cancer. The risk factors for proximal colon cancer are not well known, although general recommendations for lower risk may apply to colorectal cancer overall. This would involve a low fat intake (20 to 25 percent of calories), use of monounsaturated fats such as olive oil, an adequate fiber intake (25 to 30 g/day), moderate alcohol consumption (an average of 2 drinks/day), fish 2 to 3 times a week, an increased calcium (lowfat milk or yoghurt) intake (1,200 to 1,500 mg), increased consumption of vegetables and fruits (ideally 5 to 9 servings/day), tea (4 to 5 cups per day), and regular exercise.

Breast

Breast cancer is the third most common cancer in the world; every year about 600,000 new cases are detected, which is about 9 percent of the global can-

cer burden. It is important to distinguish between premenopausal breast cancer, in which diet plays a minor role (except for some protection afforded by consumption of vegetables and fruits, including soy products), and peri- and postmenopausal disease, in which diet may exert important controlling effects. Breast cancer occurs almost exclusively in women, and in high-risk areas (North America and western Europe), the incidence is about 4 to 30 times higher than in low-risk areas like China, Japan, and Sri Lanka, although there has been an appreciable increase in Japan during the last decade – the result of a westernization of dietary customs.

Important risk factors for breast cancer are a family history of the disease, a low number of offspring, avoidance of breast feeding of infants, a late age at first pregnancy, an early menarche, a late age at menopause, and high consumption of fats (about 30 to 40 percent of calories) and, possibly, alcohol. During the last decade, increasing evidence has been adduced indicating that there is an inverse relationship between breast cancer and increased intake of vegetables and fruits. Food antioxidants (such as selenium, retinoids, and polyphenols), as well as bran cereal fibers, have been suggested as inhibiting factors. Obviously, endocrine factors are important in breast cancer development. Fat may increase breast cancer risk by its control of hormonal regulation. In addition, high fat and high energy intakes, coupled with lack of exercise, lead to obesity, a possible contributory factor in breast cancer in postmenopausal women.

Obesity, however, seems inversely related to the risk of breast cancer in premenopausal women. F De Waard has developed a unifying concept on the etiology of breast cancer, which focuses on the events that occur during adolescence and early reproductive ages (see Weisburger and Kroes 1994). He has suggested that preneoplastic lesions develop at early ages, from 15 years onward. Several factors, such as nutritional status, high fat intake, low consumption of protective vegetables, fruit, and fibers, along with reproductive life, interact in inducing a long period of cell proliferation without sufficient differentiation in the breast.

On the other hand, early pregnancy and long-term lactation will raise the differentiation of cells, thus limiting the proliferation of less differentiated cells, the latter being more vulnerable to genotoxic attack. Fat may also influence the immune system, increase prostaglandin synthesis, and increase membrane fluidity, all phenomena bearing on the promotion and growth of neoplastic cells. Therefore, the appropriate dietary preventive measures are avoidance of heavily fried or broiled meats, a limited fat intake (possibly as low as 20 to 25 percent of total calories), preference for monounsaturated fats such as olive oil, an increased intake of vegetables, fruits, tea, and insoluble bran cereal fiber, and an energy intake that bal-

ances energy need with the avoidance of obesity. In this latter connection, an increase of exercise has been shown to lower risk and assist in weight control.

Endometrium and Ovary

Endometrial cancer strikes approximately 150,000 women in the world each year, with tenfold differences in incidence, depending on location. High incidences are found in Argentina, the United States, Canada, and western Europe, whereas a low incidence has been noted in Asian populations. Identified risk factors are, in particular, endogenous estrogen and higher amounts of exogenous hormones employed for the management of menopausal and postmenopausal symptoms. Obesity and fat consumption are also associated with increased risk. Estrogen therapy, as practiced for postmenopausal symptoms between 1960 and 1975, has been documented as a causal element for endometrial cancer, most probably because it was given in relatively large dosages and was not balanced by progesterone. If the action of limited amounts of estrogens is balanced by progesterone, cancer risk is decreased.

The role of obesity or high fat consumption in endometrial cancer may be explained by the fact that fat cells produce estrogen, which itself is a key effector in neoplastic development through its specific effect on endometrial tissue and on overall endocrine balances. As dietary factors may be responsible for an appreciable percentage of cases, limited fat intake and avoidance of excessive energy intake are suggested preventive measures. Regular exercise, likewise, constitutes a protective element.

Ovarian cancer is common in western Europe and North America, whereas it has a low frequency in Indian, Japanese, and other Asian populations. Unlike that of many other types of cancer, the incidence of ovarian cancer in the Western world has remained rather constant over time. The risk factors for ovarian cancer are the same as those for breast and uterine cancer, meaning a positive association with endocrine factors and dietary fat intake and a negative association with parity and elements that suppress ovulation. Thus, oral contraceptives may substantially reduce the risk of ovarian cancer. Limited fat intake (perhaps 20 to 25 percent of calories or less) and consumption of vegetables and fruits are suggested as preventive measures.

Pancreas

Pancreatic cancer occurs more frequently in developed countries, comprising approximately 3 percent of the worldwide cancer burden. The disease, however, is increasing in incidence over time and has a very high mortality rate because of late diagnosis and, thus, has low success in therapy. Every year, approximately 140,000 new cases are diagnosed. In the last 40 years, pancreatic cancer incidence has doubled in

western Europe and quadrupled in Japan (Hirayama 1989).

Tobacco smoking has been implicated as a major risk factor, which can explain the increasing incidence, especially in those countries where the pancreatic cancer incidence is still relatively low. Convincing evidence also exists from experimental animal research that carcinogens from tobacco and a high fat intake are positively related, whereas caloric restriction, selenium, and retinoids are inversely related. Of interest is the role shown by trypsin inhibitors in pancreatic carcinogenesis in experimental animals. These trypsin inhibitors do reduce trypsin levels in the gut, stimulating the secretion of cholecystokinin (CCK) as a feedback phenomenon. CCK stimulates pancreatic growth, thus promoting pancreatic carcinogenesis. Trypsin inhibitors, present in soy proteins, are heat labile. Soy proteins are high-quality foods, but they should be incorporated in foods and cooked (Watanapa and Williamson 1993).

Epidemiological research reveals a positive relationship for dietary fat, fried or grilled meats, and, possibly, alcohol or cholesterol, whereas an inverse relationship has been observed for caloric restriction, omega-3 fatty acids (fish and some seeds like flax seed), and fresh fruits and vegetables (Bueno de Mesquita 1992). Preventive potential has been estimated to be 70 percent. Cessation of tobacco smoking, moderate alcohol use, low fat consumption, and increased intake of vegetables and fruits are the main measures for prevention. This is particularly important because of its grim prognosis. Thus, control is optimal through prevention by lifestyle adjustment.

Prostate

Prostate cancer is the fifth most common cancer among males, and especially predominant in older males. Approximately 240,000 new cases of clinical invasive prostate cancer occur each year, and high-incidence areas are northwestern Europe and North America; in the latter, African-Americans have a particularly high incidence. Low rates are found in India, China, and Japan. There exists a 50-fold difference between populations with the highest rates of prostate cancer (blacks in Detroit, Michigan) and populations with the lowest incidence (Asians in Shanghai, China) (Nomura and Kolonel 1991). Endocrine factors may play a role in prostate carcinogenesis, but geographic pathology indicates that dietary factors are probably also important. Populations with a tradition of high fat and high protein intake have a high risk. The diet controls the endocrine balance.

Negative associations have been suggested for vitamin A, beta-carotene, vegetables, fruits, selenium, fish, and fiber. Sugar and egg consumption are weakly positive (Bosland 1986). Genetic, sexual, and dietary factors seem to play a role in prostate carcinogenesis, indicating a multifactorial process. As is true for other endocrine-controlled neoplasms, a dietary regime low

in fat and rich in vegetables and fruits, coupled with regular exercise, may contribute to lower risk irrespective of sexual and genetic elements (Wynder, Rose, and Cohen 1994).

Lung

It is surprising to note that more and more data have become available to indicate that lung cancer is influenced by dietary factors. Clearly, the disease is associated with cigarette smoking, but since E. Bjelke (1975) and G. Kvale, Bjelke, and J. J. Gart (1983) found in metabolic epidemiological studies that smokers with a higher level of vitamin A in plasma had a lower risk of lung cancer, more attention has been given to dietary factors (Ziegler et al. 1992; Le Marchand et al. 1993). Also, for humans, an inverse relationship between lung cancer development and fruit and vegetable intake has been observed, whereas other data suggest a positive relationship between dietary fat intake and lung cancer (Wynder, Taioli, and Fujita 1992). In addition, the antioxidants in tea may provide a protective effect. Currently, there are more smokers in Japan than in the United States or the United Kingdom, but the incidence of lung cancer is lower in Japan. It has been suggested that the Japanese have a lower risk because of a lower total fat intake and more frequent intake of fish, soy foods, and tea.

Therefore, although the first recommendation should be to quit smoking – or, in fact, never to start – an increased intake of fruits and vegetables (especially those containing retinoids) and also of fish, soy-derived foods, and tea, coupled with a lowered fat consumption, may serve as preventive measures and could be particularly appropriate for ex-smokers.

Food Additives, Contaminants, and Natural Toxins

For decades, the possibility of cancer risks from food additives and contaminants has been widely publicized, especially in the developed countries, where there has been an increase in the addition of various substances to food for preservative and commercial purposes. Thus, food additives and contaminants are viewed by many as a major threat to human health – and one that may cause cancer. Scientific information, however, shows exactly the opposite: Food additives are safer than everyday traditional nutrients, and the same is true for most contaminants (Miller 1992; Weisburger 1994; Weisburger and Kroes 1994).

Such opposite perceptions may be explained by the misinterpretation of epidemiological reports in the late sixties, when the term “environmental” (as in “environmental factors”) was used to account for major causes of cancer. In fact, what was meant was as lifestyle factors, but the general public (and especially the news media) misinterpreted this to mean synthetic chemicals, including food additives and contaminants.

In addition, several episodes have enhanced this misconception, as, for example, when certain food additives (that is, some food dyes in Western countries and the preservative AF-2 in Japan) were first permitted and later correctly withdrawn because of their demonstrated carcinogenicity in animals (Sugimura 1992). Regulatory action, especially in the United States, aimed at such chemicals as sodium saccharin and cyclamate, further deepened public suspicion. Yet the latter substances are now considered safe, at least at the normal intake levels that humans experience. In fact, certain substances with antioxidant properties, which are used as food additives, are even believed to reduce cancer risk. Thus, Wynder and Gori (1977), as well as Doll and Peto (1981), have estimated that cancer mortality from food additives ranges from –5 to +2 percent, the negative score specifically addressing the beneficial aspects of antioxidants used in foods.

Additives are used to improve the stability and storability of foods, as well as their flavor, consistency, appearance, texture, and nutritional quality. In certain cases, they are a necessity, such as in the case of preservatives that prevent food-borne microbial infections. And in any event, the risk of disease from food additives today is minimal, because efficient and effective control practices are available and applied to ensure safety.

Contaminants of human-made origin are, like food additives, extensively tested in animals before use, and the levels permissible in crops are well controlled internationally. Thus, the margin of safety for pesticide residues in food runs usually between 1,000 and several millions, whereas for several macro- and micronutrients, the margin of safety is as small as 2 to 10 (Kroes in press). In fact, B. N. Ames and colleagues (1990, 1992) have listed a number of naturally occurring substances in food that, because of uncontrolled exposure, provide much more concern for cancer risk than synthetic chemicals. About half of such natural chemicals that have undergone standard high-dose animal cancer tests proved to be animal carcinogens, such as the mold-generated hepatic carcinogen, aflatoxin (International Agency for Research on Cancer 1993b). In addition, as noted, powerful carcinogens are formed during the cooking of meats and during the salting and pickling of some fish and meats.

The natural defenses of humans, however, may make them capable of detoxifying low doses of most toxins, whether synthetic or natural. For example, despite a continuing low-level presence of aflatoxin B₁ in some foods, the incidence of primary liver cancer in the United States and Europe is not significant. Yet it is quite high in parts of Africa and China, where the dietary contamination is appreciable and where more people carry the hepatitis B antigen, potentiating the action.

Certainly, in light of the foregoing, it seems relevant to invest more research capacity in the identification of possible risks and benefits of naturally occurring

substances. This is especially true because many are also known to possess anticarcinogenic properties – properties that are believed to be the reason for the inverse relationship between several cancers (and heart diseases) and the regular intake of vegetables, fruits, and tea.

Food preparation has entailed cancer risk in the past and will continue to do so in the future. Preservation methods, for example, such as the use of salt or pickling solutions, are associated with a high risk of stomach cancer and in some areas, such as China, with cancer of the esophagus. Salted fish causes nasopharyngeal cancer, and salt and high nitrate (salt-peter) concentrations in several meat products can lead to the formation of carcinogenic nitroso compounds, or of the chloro analog of methionine, either in the food itself or in the stomach. Salt is cytotoxic to the gastric mucosa, translated by increased cell duplication rates and, in turn, to more efficient carcinogenesis. Some salted, pickled foods contain direct-acting mutagens thought to be gastric carcinogens (Weisburger 1992; Chen et al. 1995). Salt, not balanced by potassium from vegetables, and calcium from dairy products is also a cause of hypertension and stroke. In Japan (Sugimura 1992) and in Belgium (Joossens, Hill, and Geboers 1985), formal plans were introduced to lower salt intake by people.

Charcoal-broiled meats or fish have at their surface polycyclic aromatic hydrocarbons that are established animal carcinogens. But it is important to note that the ordinary cooking (broiling, frying) of meats or fish can produce powerful mutagens, consisting of about 19 heterocyclic amines (also established animal carcinogens) for specific target organs. They are believed to be the key carcinogens causing increased incidence of several human cancers, such as those in the breast, prostate, colon, and pancreas. Certainly it has been shown that those who generally eat well-done meat increase their risk of colon cancer. The formation of heterocyclic amines during the heating of meats can be reduced by preliminary brief microwave cooking (removing essential creatinine) or by the addition of antioxidants, soy protein, or the indole amino acids tryptophan and proline, which all compete with creatinine in the so-called Maillard reaction, forming heterocyclic amines (Weisburger and Kroes 1994).

Prevention: An Integrated Approach

A substantial amount of solid epidemiological and experimental evidence indicates that the majority of human cancers, and indeed many other chronic diseases, such as heart disease, hypertension, and adult-onset diabetes, are largely preventable. Complex causes have been, or are being, identified, and the underlying mechanisms elucidated. Control of many major diseases of humankind in the past, such as scurvy, pellagra, rickets, polio, smallpox, rabies, and

tuberculosis, has been achieved by prevention strategies. Therefore, a clear and balanced prevention approach to the effective control of human cancers (and other chronic illnesses, such as cardiovascular diseases) is likely to be successful as well.

Experience in the past two decades with other chronic ailments, such as cardiovascular diseases, indicates that the application of sometimes even simple measures can have a considerable impact on their outcome (Meyskens 1992). In fact, cancer prevention runs a decade behind the scientific understanding of the disease, as a number of lifestyle-associated factors contributing significantly to cancer risk are well known. Tobacco and nutritional traditions, in particular, and – to a lesser extent – radiation, some chemicals, and certain viruses, are documented, avoidable risk factors.

Wynder and Gori (1977), Doll and Peto (1981), and J. H. Weisburger (1992) have contributed substantially to the evidence of preventive potential for many types of cancer through their listing and documentation of avoidable risks of cancer. Unfortunately, not only are such concrete factors as tobacco, diet, lifestyle, and radiation contributing to the cancer burden, but poverty does as well, because the need for essential lifestyle changes has been difficult to communicate effectively to the lower socioeconomic groups (Tomatis 1992).

Cancer prevention programs should be based on reliable epidemiological and laboratory evidence, and on ethical and moral responsibility, and ought to specify clearly achievable outcomes in mortality and morbidity reduction. Moreover, they should be integrated with other chronic-disease-prevention programs. The multifactorial elements, such as nutritional traditions, sedentary habits, and tobacco use, which represent a risk for diseases like coronary heart disease, hypertension, stroke, obesity, and many neoplastic diseases, have to do with lifestyle. In fact, these factors are the *major causes and modulators of these diseases*.

Table IV.F2.1 depicts several realistic actions to take in order to lower the risk for certain common diseases. In the Western world, nutritional traditions with a relatively high fat intake (38 to 46 percent of energy intake), low cereal fiber, and low vegetable and fruit consumption, along with a lack of regular physical exercise, are associated with high incidence and high cost of the management of chronic diseases, such as cardiovascular diseases, diabetes, obesity, and specific types of cancer. In the Far East and in Central and South America, prevailing illnesses seem to stem from other nutritional traditions, such as the use of highly salted and pickled foods and a limited variety in diet. Currently, the changing nutritional habits in Japan to a Western style parallel an increase in heart disease and the kind of cancers common in the Western world. This provides strong additional support for the thesis that dietary customs and specific chronic diseases are related.

Today's knowledge enables us to recommend a healthy dietary regime in which fats should be

Table IV.F.2.1. *Chronic disease prevention and health promotion*

Action	Lower risk
1. Quit smoking or use a low-tar and nicotine cigarette.	Coronary heart disease; cancers of the lung, oropharynx, kidney, bladder, pancreas, cervix
2. Quit chewing tobacco or snuff dipping.	Cancer of oral cavity, esophagus
3. Lower the fat intake to at least 20–30% of energy (calorie) intake, preferably lower (15–20%).	Coronary heart disease; cancers of the distal colon, postmenopausal breast, prostate, pancreas, ovary, endometrium
4. Limit protein intake to 10–15% of calories; include vegetable proteins.	Colon cancer, kidney diseases, aging
5. Increase vegetables and fruits considerably (ideally 5–9 servings); provide sufficient essential micronutrients.	Most chronic diseases, aging
6. Increase fiber, soluble as well as insoluble (1:1), to 25–35 grams.	Cancer of colon, breast; constipation, diverticulosis, appendicitis
7. Increase Ca ²⁺ intake (skim milk, yoghurt, some vegetables, sardines).	Cancer of colon, osteoporosis
8. Reduce intake of heavily fried or grilled foods, especially meats and gravies.	Cancers of colon, breast, pancreas; heart disease
9. Avoid pickled, smoked, and salted foods.	Cancer of stomach, esophagus, nasopharynx, liver; hypertension, stroke
10. Lower salt Na ⁺ intake; balance K ⁺ + Ca ⁺⁺ /Na ⁺ ratio.	Hypertension, stroke, cardiovascular disease
11. Have regular moderate exercise.	Coronary heart disease; cancer of the colon, breast; aging
12. Maintain proper weight; avoid obesity.	Coronary heart disease; cancer of the endometrium and kidneys; postmenopausal breast cancer; diabetes
13. Moderate use of alcoholic beverages.	Liver cirrhosis, liver cancer, cardiomyopathy, male impotence, oropharyngeal and esophageal cancer
14. Practice sexual hygiene.	Cancer of the cervix, penis; AIDS
15. Moderate sun exposure.	Skin cancer, melanoma, skin aging

replaced by complex carbohydrates (starches) that should provide around 70 to 75 percent of the calories needed for energy. Furthermore, a protein intake of between 10 to 15 percent from animal and vegetable sources (more in growing young children, less in older individuals) is recommended. Fats should consist of a fair proportion of monounsaturated oils, such as olive oil or canola oil, and omega-3 fatty acids as found in fish and some seeds, like flaxseed.

Excessive salting, pickling, and smoking of food as a mode of preservation should, ideally, be abandoned, and foods should be preserved by refrigeration or freezing, or eaten fresh. Caloric intake should be equal to energy need as an effective means to avoid obesity. The intake of fruits and vegetables should be increased considerably – ideally to more than 5 servings per day. Bran cereal fibers, or breads baked with high-fiber flour, increase stool bulk, avoid constipation, and lower the risk of colon and breast cancer and perhaps other diseases. A low intake of total salt, 3g/day or less, and adequate calcium (1,000 to 1,500 mg) and magnesium (300 mg) are beneficial; less is needed on a low-protein diet. Moderate, but regular, physical exercise is also part of a healthy lifestyle. Alcohol consumption ought to be moderate, but adequate fluid intake (2 to 2.5 liters daily for adults) is essential for maintenance of physiologic functions. Tea, an extract of the plant *Camellia sinensis*, is, after water, the second-most-used beverage in the world. Because it is made with boiling water, it is sterile even if the water source is not pure. Tea is rich in fluoride, potassium, and espe-

cially in antioxidants that lower the risk of coronary heart disease and many types of cancer (Weisburger 1996; Weisburger and Comer this volume).

These recommendations will contribute to better health by lowering the risk for major chronic diseases. In education and in medical practice, emphasis must be placed on the importance and the efficacy of available methods for chronic disease prevention. One task of practicing nutritionists and home economists ought to be that of devising practical recipes for cooks to incorporate the essence of a new, health-promoting lifestyle that the public will find attractive. Indeed, it is essential to devise appealing dishes and drinks for the public that are also designed for chronic disease prevention, and when such preventive approaches are successful, health-care costs should decrease. Good health into old age is not only a desirable goal for the individual but may have major economic savings for the population at large. To repeat our earlier quotation, the ultimate goal, as expressed by Ernst Wynder, should be “to die young, as late in life as possible” (Wynder et al. 1994).

Robert Kroes
J. H. Weisburger

Research in Weisburger's laboratory is supported by USPHS-NIH grants and contracts from the National Cancer Institute, grants from the Tea Trade Health Research Association and the American Cancer Society, and gifts from Texaco, Inc., and the Friends Against Cancer Team.

Note

1. For ease of reading, only specific references have been inserted. Our discussion has been based mainly on these publications: National Research Council 1982; Dutch Nutrition Council 1986; Kroes 1986; Reddy and Cohen 1986; U.S. Public Health Service 1988; Committee on Diet and Health, Food and Nutrition Board 1989; International Agency for Research on Cancer 1990, 1993a; Weisburger 1991, 1992; Weisburger and Kroes 1994; Levi, Lucchini, and La Vecchia 1994.

Bibliography

- Adamson, R. H., J. A. Gustafsson, N. Ito, et al. 1995. Heterocyclic amines in cooked foods. *Proceedings of the 23rd International Symposium of the Princess Takamatsu Cancer Research Fund*. Princeton, N.J.
- Ames, B. N., and L. S. Gold. 1992. Animal cancer tests and cancer prevention. *Journal of the National Cancer Institute Monographs*, no. 12: 125–32.
- Ames, B. N., M. Profet, and L. S. Gold. 1990. Nature's chemicals and synthetic chemicals: Comparative toxicology. *Proceedings of the National Academy of Sciences* 87: 7782–6.
- Baron, J. A., M. Gerhardson de Verdier, and A. Ekblom. 1994. Coffee, tea, tobacco, and cancer of the large bowel. *Cancer Epidemiology, Biomarkers, Prevention* 3: 565–70.
- Bjelke, E. 1975. Dietary vitamin A and human lung cancer. *International Journal of Cancer* 15: 561–5.
- Block, G., B. Patterson, and A. Subar. 1992. Fruit, vegetables and cancer prevention: A review of the epidemiological evidence. *Nutrition and Cancer* 18: 1–29.
- Bosland, M. C. 1986. Diet and cancer of the prostate: Epidemiologic and experimental evidence. In *Diet, nutrition, and cancer: A critical evaluation*, Vol. 1, ed. B. S. Reddy and L. A. Cohen, 125–49. Boca Raton, Fla.
- Brugge, J., T. Cuppan, E. Harlow, and F. McCormick. 1991. *Origins of human cancer; a comprehensive review*. Plainview, N.Y.
- Bueno de Mesquita, H. B. 1992. On the causation of cancer of the exocrine pancreas: a population based case-control study in the Netherlands. Thesis, University of Utrecht.
- Castelletto, R., X. Castellsague, M. Muñoz, et al. 1994. Alcohol, tobacco, mate drinking and esophageal cancer in Argentina. *Cancer Epidemiology, Biomarkers, Prevention* 3: 357–64.
- Chen, W., J. H. Weisburger, E. S. Fiala, et al. 1995. 2-Chloro-4-Methylthiobutanoic acid, a newly identified mutagen in salt, nitrite-treated fish. *Nature* 374: 599.
- Cohen, S. M., and L. B. Ellwein. 1992. Risk assessment based on high dose animal exposure experiments. *Chemical Research in Toxicology* 5: 742–8.
- Committee on Diet and Health, Food and Nutrition Board. 1989. *Diet and health: Implication for reducing chronic disease risk*. Washington, D.C.
- Correa, P. 1992. Human gastric carcinogenesis: A multistep and multifactorial process. First American Cancer Society Award Lecture on Cancer Epidemiology and Prevention. *Cancer Research* 52: 6735–40.
- Craddock, V. M. 1992. Aetiology of oesophageal cancer: Some operative factors. *European Journal of Cancer Prevention* 1: 89–103.
- Davidson, M. K., I. K. Davis, and J. R. Lindsey. 1987. Requirement and selection of an animal model. *International Journal of Medical Sciences* 23: 551–5.
- Doll R., and R. Peto. 1981. The causes of cancer: Quantitative estimate of avoidable risks of cancer in the United States today. *Journal of The National Cancer Institute* 66: 1191–308.
- Dutch Nutrition Council. 1986. Factoren in de voeding en het ontstaan van kanker (Dietary factors in relation to the causation of cancer). *Dutch Nutrition Council Report*. The Hague.
- Echols, M., and M. F. Goodman. 1991. Fidelity mechanisms in DNA replication. *Annual Review of Biochemistry* 60: 477–511.
- Galloway, D. J. 1989. Animal models in the study of colorectal cancer. *Cancer Surveys* 8: 170–88.
- Harris, E. D. 1991. Regulation of anti-oxidant enzymes. *Federation of American Societies for Experimental Biology Journal* 6: 2675–83.
- Highman, B., M. J. Norvell, and T. E. Shellenberg. 1977. Pathological changes in female C₃H mice continuously fed diets containing diethylstilbestrol or 17-beta-estradiol. *Journal of Environmental Pathology and Toxicology* 1: 1–30.
- Hirayama, T. 1989. Epidemiology of pancreatic cancer in Japan. *Japanese Journal of Clinical Oncology* 19: 208–15.
- Howson, C. P., T. Hiyama, and E. L. Wynder. 1986. Decline of gastric cancer: Epidemiology of an unplanned triumph. *Epidemiology Reviews* 8: 1–27.
- International Agency for Research on Cancer. 1990. Cancer: Causes, occurrence and control. *International Agency for Research on Cancer Scientific Publication no. 100*. Lyon.
1993. Some naturally occurring substances, food items and constituents, heterocyclic aromatic amines and mycotoxins. *Monographs on the evaluation of carcinogenic risk to humans*, no. 56. Lyon.
- Joossens, J. V., M. J. Hill, and J. Geboers, eds. 1985. *Diet and human carcinogenesis*. Amsterdam.
- Kroes, R. 1979. Animal data, interpretation and consequences. In *Environmental carcinogenesis*, ed. P. Emmelot and E. Kriek, 287–302. Amsterdam.
1986. Diet in relation to cancer. In *New concepts and developments in toxicology*, ed. P. L. Chambers, P. Gehring, and F. Sakai, 91–105. New York.
1987. Contribution of toxicology towards risk assessment of carcinogens. *Archives of Toxicology* 60: 224–8.
- In press. *Risk assessment: Past, present and future*. Proceedings of the 2d Princess Chulabhorn Science Congress. Thailand.
- Kroes, R., R. B. Beems, M. C. Bosland, et al. 1986. Nutritional factors in lung, colon and prostate carcinogenesis in animal models. *Federation Proceedings* 45: 136–41.
- Kvale, G., E. Bjelke, and J. J. Gart. 1983. Dietary habits and lung cancer risk. *International Journal of Cancer* 31: 397–405.
- Land, C. E., J. D. Boice, C. E. Shore, et al. 1980. Breast cancer risk from low dose exposure to ionizing radiation: Results of parallel analysis of three exposed populations of women. *Journal of the National Cancer Institute* 65: 353–76.
- Le Marchand, L., J. H. Hankin, L. N. Kolonel, et al. 1993. Intake of specific carotenoids and lung cancer risk. *Cancer Epidemiology, Biomarkers, Prevention* 2: 183–7.
- Levi, F., F. Lucchini, and C. La Vecchia. 1994. Worldwide patterns of cancer mortality, 1985–1989. *European Journal of Cancer Prevention* 3: 109–43.
- Liehr, J. G. 1990. Genotoxic effects of estrogens. *Mutation Research* 238: 269–76.
- Meyskens, F. L. 1992. Strategies for prevention of cancer in humans. *Oncology* 6 (supplement): 15–24.
- Miller, S. A. 1992. Food additives and contaminants. In *Toxicology: The basic science of poisons*. Fourth edition, ed. L. J. Casarett and J. Doull, 819–53. New York.

1994. A brief history of chemical carcinogenesis. *Cancer Letters* 83: 9-14.
- National Research Council. 1982. *Diet, nutrition, and cancer*. Washington, D.C.
- Negri, E., C. La Vecchia, S. Franceschi, and A. Tavani. 1993. Attributable risk for oral cancer in northern Italy. *Cancer Epidemiology, Biomarkers, Prevention* 2: 189-93.
- Nomura, A. M. Y., and L. N. Kolonel. 1991. Prostate cancer: A current perspective. *American Journal of Epidemiology* 133: 200-27.
- Olson, J. A., and S. Kobayashi. 1992. Antioxidants in health and disease: An overview. *Proceedings of the Society for Experimental Biology and Medicine* 200: 245-7.
- Packer, L. 1992. Interactions among antioxidants in health and disease: Vitamin E and its redox cycle (43433). *Proceedings of the Society for Experimental Biology and Medicine* 200: 271-6.
- Parkin, D. M. 1993. Studies of cancer in migrant populations. In *Cancer in Italian migrant populations*, eds. M. Gaddels, D. M. Parkin, and M. Khlat et al. Lyon: International Agency for Research on Cancer, Scientific Publication no. 123, 1-10.
- Randerath, K., E. Randerath, T. F. Danna, et al. 1989. A new sensitive ^{32}P -postlabeling assay based on the specific enzymatic conversion of bulky DNA lesions to radiolabeled dinucleotides and nucleoside 5'-monophosphates. *Carcinogenesis* 10: 1231-9.
- Reddy, B. S. 1992. Animal experimental evidence on macronutrients and cancer. In *Macronutrients: Investigating their role in cancer*, ed. M. S. Micozzi and T. M. Moon, 33-54. New York.
- Reddy, B. S., and L. A. Cohen. 1986. Diet, nutrition, and cancer: A critical evaluation. In *Macronutrients and cancer*, Vol. 1. Boca Raton, Fla.
- Ronai, Z. 1992. Ras oncogene detection in preneoplastic lesions: Possible applications for diagnosis and prevention (Review). *Oncology Research* 4: 45-8.
- Sai, K., T. Umemura, A. Tagaji, et al. 1992. The protective role of glutathione, cysteine and vitamin C against oxidative DNA damage induced in rat kidney by potassium bromate. *Japanese Journal of Cancer Research* 83: 45-51.
- Seitz, H., and U. A. Simanowski. 1991. Alcohol and cancer: A critical review. In *Alcoholism: A molecular perspective*, ed. T. N. Palmer, 275-96. New York.
- Sugimura, T. 1992. Multistep carcinogenesis: A 1992 perspective. *Science* 258: 603-7.
- Tannenbaum, A. 1942a. The genesis and growth of tumors: II. Effects of caloric restriction per se. *Cancer Research* 2: 460-7.
- 1942b. The genesis and growth of tumors: III. Effect of a high fat diet. *Cancer Research* 2: 468-74.
- Tannenbaum, A., and H. Silverstone. 1953. Nutrition in relation to cancer. *Advances in Cancer Research* 1: 453.
- Tomatis, L. 1992. Poverty and cancer. *Cancer Epidemiology, Biomarkers, Prevention* 1: 167-75.
- Trosko, J. E., and J. I. Goodman. 1994. Intercellular communication may facilitate apoptosis: Implications for tumor promotion (Review). *Molecular Carcinogenesis* 11: 8-12.
- U.S. Public Health Service. 1988. *The Surgeon General's Report on Nutrition and Health*, Department of Health and Human Services (Public Health Service) Publication no. 88-50210. Washington, D.C.
- Watanapa, P., and R. C. N. Williamson. 1993. Experimental pancreatic hyperplasia and neoplasia: Effects of dietary and surgical manipulation. *British Journal of Cancer* 67: 877-84.
- Weisburger, J. H. 1991. Nutritional approach to cancer prevention with emphasis on vitamins, antioxidants and carcinogens. *American Journal of Clinical Nutrition* 53: 226-37.
1992. Mechanisms of macronutrient carcinogenesis. In *Macronutrients: Investigating their role in cancer*, ed. M. Dekker, M. S. Micozzi, and T. E. Moon, 3-31. New York.
1994. Commentary. Does the Delaney Clause of the U.S. Food and Drug Laws prevent human cancers? *Fundamental and Applied Toxicology* 22: 483-93.
1996. Tea antioxidants and health. In *Handbook of Antioxidants*, eds. Enrique Cajenas and Lester Packer. New York. 469-86.
- Weisburger, J. H., and R. Kroes. 1994. Mechanisms in nutrition and cancer. Meeting report. *European Journal of Cancer Prevention* 3: 293-8.
- Weisburger, J. H., and G. Williams. 1991. Critical effective methods to detect genotoxic carcinogens and neoplasm promoting agents. *Environmental Health Perspectives* 90: 121-6.
- Williams, G. M., and J. H. Weisburger. 1991. Chemical carcinogens. In *Casarett and Doull's toxicology*. Fourth edition, ed. M. V. Amdur, J. Doull, and C. D. Klaassen, 127-200. New York.
- World Cancer Research Fund. 1997. *Food, nutrition, and the prevention of cancer: A global perspective*. Washington, D.C.
- Wynder, E. L., and G. B. Gori. 1977. Contribution of the environment to cancer incidence: An epidemiologic exercise. *Journal of the National Cancer Institute* 58: 825-32.
- Wynder, E. L., J. A. Peters, and S. Vivona, eds. 1975. Nutrition in the causation of cancer, a symposium. *Cancer Research* 35: 3231-35.
- Wynder, E. L., O. P. Rose, and L. A. Cohen. 1994. Nutrition and prostate cancer: A proposal for dietary intervention. *Nutrition and Cancer* 22: 1-10.
- Wynder, E. L., E. Taioli, and Y. Fujita. 1992. Ecologic study of lung cancer risk factors in the U.S. and Japan with special reference to smoking and diet. *Japanese Cancer Research* 83: 418-23.
- Yamasaki, H. 1990. Gap junctional intercellular communication and carcinogenesis. *Carcinogenesis* 11: 1051-8.
- Zeng, Y., H. Oshima, G. Bouvier, et al. 1993. Urinary excretion of nitrosamino acids and nitrite by inhabitants of high and low risk areas for nasopharyngeal carcinoma in southern China. *Cancer Epidemiology, Biomarkers, Prevention* 2: 195-200.
- Ziegler, R. G., A. F. Subar, N. E. Craft, et al. 1992. Does beta-carotene explain why reduced cancer risk is associated with vegetable and fruit intake? *Cancer Research* 52: 2060-6.

IV.F.3 Nutrition and Heart-Related Diseases

...When we have stuffed
These pipes and those conveyances of our
blood
With wine and feeding ...

(Allbutt 1915: 239, citing Traube)

Over the course of the twentieth century, cardiovascular disease (CVD) has become the leading cause of death in the United States. CVD is also a significant

cause of morbidity and mortality in many other industrialized countries and regions, such as Scandinavia, the United Kingdom, Australia, and Canada. Most CVD is manifested as coronary artery disease (CAD), usually based on atherosclerosis. This “epidemic” of CVD has been attributed to the poor lifestyle habits of members of late-twentieth-century industrialized, urban society, who smoke tobacco, exercise rarely, and indulge in fat-laden diets (Kannel 1987).

A striking similarity of these factors leading to disease is that each – in most cases – can be modified by an individual at risk for coronary disease, even without professional guidance and in the absence of public health initiatives. But risk factors are not always easily eliminated. Addiction to tobacco is difficult to overcome. Exercise may be problematic for some people, given constraints on time posed by other obligations. Everyone, however, must eat, and perhaps for this reason, of all the possible causes of heart-related diseases, diet has received the most attention.

This chapter explores the relationship between nutrition and heart-related diseases by describing selected nutrients that have been implicated in the pathogenesis, prevention, or treatment of CAD. Most of the available data come from population studies. It appears unlikely that any single nutrient will soon be identified as the specific dietary agent that causes atherosclerotic diseases. Moreover, any individual nutrient is but a small part of a larger group of chemicals that make up any particular “food.” As of this writing, the overall relationship between diet and heart disease remains obscure. It will require many years of medical, nutritional, and food science research before we can hope to weave into a meaningful tapestry all of the puzzling threads of nutritional biochemistry and CAD pathophysiology.

Pathogenesis of Atherosclerosis

In the middle of the nineteenth century, it was discovered that lipids constituted a major component of atherosclerotic plaque (Vogel 1847). This observation led investigators to wonder how this lipid collection was formed and how it might be related to disease. They noted that when people with CAD died, examination of the vessels that supply blood to the heart (the coronary arterial endothelium) revealed greasy, yellow “plaques” adhering to the endothelial cells. These plaques varied in thickness and composition, reflecting both the age of the subjects and the severity of their disease. A few decades into the twentieth century it was widely recognized that CAD lesions were due to atherosclerosis, with cell death within the lesion (Leary 1935). Additional clues to disease severity – and the longevity of the disease process – were later found in the presence of blood substances, such as lipids and macrophages, and in structural damage to the surrounding endothelium.

Noting precisely which endothelium is diseased

may provide clues to the disease process. Although CAD is often conceptualized as a systemic disease, the location of atherosclerotic lesions is not random, nor are they necessarily “systemic.” Some areas of the blood vessels are more “lesion-prone” than others, and these sections differ structurally and functionally from “non-lesion-prone” sites (Schwartz, Valente, and Sprague 1993). Lesion-prone sites are more permeable to substances that lead to the development of CAD. Any adequate explanation of the disease process must explain these morphological and topographical features.

Although autopsy findings may serve to document an underlying disease process, they represent only one moment in the lifetime of the affected individual and cannot alone explain the natural history of the disease. One approach to understanding the disease process is to conduct experiments using an animal that exhibits a similar set of lesions. In 1943, D. V. Dawber and L. N. Katz studied chickens that developed atheromata and suggested that their findings might be relevant to spontaneous disease (Dawber and Katz 1943).

Pigs also provide a useful model for studying human atherosclerotic CAD – a model used to good effect by investigators in the 1960s and early 1970s. When pigs were fed a low-fat, cholesterol-free diet, researchers initially noted a degeneration of smooth muscle cells. As the pigs aged, the degenerated smooth muscle cells accumulated and caused intimal thickening. If there was no injury to the blood vessel during this thickening stage, the pigs experienced no disruption of the endothelium of the artery. However, if the investigators mechanically injured the endothelium, intimal thickening progressed to form a more substantial plaque – one that could impede blood flow within the artery.

This early plaque contained not only degenerated smooth muscle cells but also collagen, a fibrous tissue that made the plaque less compliant. Plaque thickening could be further accelerated (in concert with mechanical injury) with dietary factors, including oxidized sterols, such as vitamin D₃ and 25-hydroxy cholesterol (a derivative of pure cholesterol). In advanced stages of atherosclerosis, lipid infiltrated the plaque. Other factors were found to influence development of the lesion. As the vessel lumen narrowed during plaque development, reduced blood flow through the area reduced oxygen delivery to the structures supplied by that blood vessel, and the hypoxemia thus produced sped up lipid accumulation in the plaque. As the plaque continued to thicken, the endothelium was stretched thin, sometimes thin enough to rupture.

The gap left by this process was then vulnerable to infiltration by lipoproteins. In animals that did not have the intimal injury, there was no progression past the intimal thickening. When investigators used electron microscopy to compare the experimentally induced atherosclerotic vessels of the mature pig

with those obtained from a human who had undergone open heart surgery for CAD, they could detect no histological differences.

These animal studies were vital to our present understanding of the atherosclerotic process in humans. Recent research supports the concept that the arterial endothelium must sustain a mechanical injury to begin the process of pathologically significant plaque formation. Other blood components now known to be involved in this process are monocytes, platelets, and lipid-protein complexes, such as low-density lipoprotein (LDL) and lipoprotein(a). Also part of the process of plaque formation are chemical mediators, such as adhesive cytokines, chemoattractants, free radicals, and proteolytic enzymes.

Currently, most investigators believe that in early atherogenesis, monocytes (a type of white blood cell) are recruited to a "lesion-prone" area of the artery's innermost wall, or intima. Before the plaque can be formed, the monocyte has to attach to the endothelium, a process orchestrated by various chemoattractant substances (for example, oxidatively modified LDL cholesterol) and adhesive cytokines (for example, interleukin 1-beta). Once attached, the monocytes migrate through the endothelium and attach to its underside, away from the portion of the endothelium that is in contact with the bloodstream.

In this new space, the monocytes are transformed into a different type of cell, known as macrophages. As long ago as the early 1900s, pathologists observed macrophages associated with mature atherosclerotic plaques, but their presence was believed to be incidental. In the late twentieth century, however, macrophages have been assigned a pivotal role in atherogenesis. The macrophage synthesizes a variety of substances that activate the inflammatory response in the affected area, such as oxygen free radicals, proteases, and lipases. These cause the macrophage to take up oxidatively modified LDL and change its appearance so that the cell has a foamy appearance when viewed with a microscope. The cells are thus called "foam cells."

The next step in the process of atherosclerosis is necrosis of the foam cell, likely due to cytotoxicity from the oxidatively modified LDL. Smooth muscle migration and proliferation are mediated by a platelet-derived growth factor, a potent chemoattractant. Fibroblastic growth factors probably regulate smooth muscle cell proliferation. In the area of the plaque where there is the greatest density of macrophages, rupture may occur due to the high local concentration of macrophage-derived metalloproteases. This process ultimately results in blood clots that can obstruct blood flow – mural or occlusive thrombosis. Occlusion, whether temporary or not, creates a locally hypoxemic environment, which further enhances plaque growth. Oxidized LDL may initiate an autoimmune process that also adds to the inflammatory reaction occurring in the lesion-prone area.

The intricate atherosclerotic process provides investigators with many avenues of exploration for prevention and intervention. From a nutritional standpoint, the most obvious factors are cholesterol, dietary fat, and the biochemical precursors for the substances that may enhance or intervene in plaque formation: omega-3 fatty acids, protein, "antioxidant" nutrients (for example, vitamin E, beta-carotene, vitamin C), and other dietary substances. As an alternative approach to dietary prevention, one might consider dietary changes that would decrease the activities of the substances that enhance arteriosclerosis.

Cholesterol

Early Epidemiological Studies

In the mid-1940s, Ancel Keys (1963) and William Kannel (1971) and their colleagues designed two large-scale studies to examine longitudinally what caused people to develop coronary artery disease. These investigators believed that by recording the essential characteristics of a large group of at-risk people and by comparing those who went on to develop CAD with those who did not, they could identify particular risk factors for the disease. These two sets of studies have had a profound influence on the development of the field.

Keys and his associates studied 281 Minnesota business and professional men who were 45 to 55 years of age and clinically healthy at the start of a 15-year study period. Each year, the men were given a detailed physical examination, with particular attention to the cardiovascular system. Examiners noted each man's weight, relative body fatness, blood pressure, and serum cholesterol concentration. They measured two different pools of cholesterol in addition to total cholesterol: high-density lipoprotein (HDL) and low-density lipoprotein (LDL) cholesterol. During the 15 study years, 32 deaths occurred, 17 of which were directly attributable to CAD. Body weight and fatness were not predictive of disease, but the men in the upper part of the blood pressure distribution had a greater risk of CAD. Most compelling were the findings with regard to cholesterol. They indicated a direct relationship between blood cholesterol and risk of CAD, extending over all levels. The men who remained healthy had significantly higher HDL cholesterol than did subjects who developed CAD (mean 45.29 milligrams [mg] per deciliter as compared with 39.45). Detailed statistical analysis, however, failed to support the hypothesis that HDL cholesterol level was an independent predictor of CAD risk. Rather, high LDL and total cholesterol were more important predictors of CVD.

The Framingham Study (Kannel et al. 1971) followed 2,282 men and 2,845 women in Framingham, Massachusetts, for 14 years, starting in 1948. As in the Keys study, investigators were attempting to identify factors related to the onset of clinical coronary dis-

ease. Subjects were free of known coronary disease at the time of enrollment and were subdivided into groups based on their serum lipid content. After 14 years, 14 percent of the men and 6 percent of the women had developed some clinical manifestation of CAD. The incidence of CAD increased with age, and the baseline serum lipids and lipoproteins were higher in the CAD group than in the other subjects. Furthermore, the lipid profiles in the CAD subjects were high when compared to profiles in other parts of the world, such as France and Japan, where low CAD rates have been reported.

Investigators were unable to demonstrate that a particular lipid played a greater role than others. However, they did observe differences between men and women. In men and younger women (<55 years), the differences between those with CAD and those without the disease were more related to the total cholesterol. In older women (>55 years), prebeta lipoprotein (very low density lipoprotein or VLDL) discriminated better than total cholesterol between women with and without CAD. These two studies, coupled with what was known about atherosclerotic plaque composition, further supported the idea that one key to solving the mystery of atherosclerosis was to determine the relationship between blood lipids and plaque formation. Moreover, of all blood lipids, cholesterol appeared to be the most important.

Sources of Cholesterol

Cholesterol occurs naturally in eukaryotic cells (cells with a membrane-bound nucleus). In humans, cholesterol serves many vital functions. It acts as an integral part of the cellular membrane, serves as a chemical backbone for essential substances (for example, steroid hormones, vitamin D), and assists in digestion through its role in the formation of bile salts. Given the central role of cholesterol in biochemical and physiological functions, it is not surprising that the substance became a central focus for biochemical researchers. Although all human cholesterol was once thought to be ingested (Leary 1935), we now know that humans are able to manufacture cholesterol *de novo*, primarily in the liver and, to a lesser extent, in intestinal and other cells. Thus, even if one consumed a very low-cholesterol diet, the adult liver and intestine would still manufacture approximately 800 milligrams of cholesterol per day, which is enough for normal human functions. But cholesterol also reaches the body through dietary means for most humans, except those who are strict vegetarians and consume neither flesh nor dairy or egg products. Thus, cholesterol in atherosclerotic plaque formation could come from either *de novo* synthesis or dietary sources. The interaction between *de novo* cholesterol synthesis and dietary cholesterol, and the subsequent metabolism of cholesterol, is an intricate and fascinating phenomenon.

Serum cholesterol is transported bound to protein,

such as apoprotein, along with phospholipids and other circulating fat-soluble compounds. These lipoproteins (or lipid:apoprotein molecules) are classified according to increasing density: From lower to higher density they are called chylomicrons, very low density lipoproteins (VLDL), low-density lipoproteins (LDL), and high-density lipoproteins (HDL). LDL typically contains 60 to 70 percent of the total serum cholesterol and HDL about 15 to 20 percent. The remaining cholesterol is carried in VLDL and chylomicrons.

The apoprotein portions contain biochemical signals that regulate the entry and exit of particular lipids at specific targets. For instance, chylomicrons transport dietary cholesterol, triglycerides, and other lipids from the intestine to the liver and adipose tissue. VLDL transports *de novo* synthesized cholesterol and triglycerides from the liver to adipose tissue. The residue is transformed into LDL, which is very rich in cholesterol. LDL moves cholesterol to peripheral tissues and regulates *de novo* cholesterol synthesis. Non-hepatic target cells possess specific LDL receptors that allow the cell to take up LDL cholesterol. The receptor number can be increased and decreased depending on the needs of the target cell. HDL presumably carries cholesterol from peripheral tissues back to the liver.

LDL regulates cholesterol metabolism through feedback inhibition. First, the cholesterol that is released from the LDL after it is inside the cell suppresses the key synthetic enzyme, HMG-CoA reductase, slowing *de novo* cholesterol formation. Second, when cholesterol concentration within a cell is adequate for its needs, the cell shuts down the manufacture of additional LDL receptors, thus preventing it from taking up more cholesterol.

Relationship of Cholesterol to CAD

Perhaps the clearest link between cholesterol, LDL, and arteriosclerosis comes from people with homozygous or heterozygous familial hypercholesterolemia. People who inherit both genes of this autosomal recessive disorder (homozygotes) have extraordinarily high levels of total and LDL cholesterol in circulation. Depending on the type of inheritance, they suffer from premature CAD either as a child or as a young adult. The defect in most cases is an absence or deficiency of functional LDL receptors, impeding the movement of LDL-cholesterol into target cells. Because LDL cannot satisfactorily get cholesterol inside a cell, there is no mechanism to turn off HMG-CoA reductase, and cholesterol biosynthesis continues unchecked. The excess cholesterol, primarily found with LDL, is deposited in various tissues, including arterial endothelium. These cause myocardial infarctions at a very young age. Because dietary cholesterol restriction does not affect *de novo* synthesis, it is of very limited value alone in the management of these individuals.

Heterozygous familial hypercholesterolemia is rela-

tively common (1:500 births), but is certainly far less common than CAD. Nonetheless, the association between cholesterol and CAD at extreme levels supports the idea that cholesterol is a factor in the development of CAD, particularly as it is consistent with numerous epidemiological studies that demonstrate a relationship between total and LDL cholesterol and CVD. Some of these are within-population studies, such as the Framingham and Minnesota studies discussed in the section "Early Epidemiological Studies." Others compare different populations or people who move from one population to another. These studies often consider the importance of culture – particularly as manifested in what people eat – and, thus, bring us closer to the overall topic of this chapter and this work, which is diet.

The observation that different communities have different rates of CAD provides an important counterbalance to overly deterministic theories of disease causation, as well as engendering a series of ideas leading toward an understanding of the possible impact of diet on cholesterol and on CAD. C. D. De Langen, a Dutch public health physician, noted in the 1910s and 1920s that the incidence of CAD was extraordinarily low among residents of the island of Java (Snapper 1963). He explained part of this difference as being the result of differences in diet; Javanese stewards who worked on Dutch steamships soon developed a pattern of coronary health similar to that of the native-born Dutchmen who worked on the ships.

In 1941, I. Snapper made a similar series of observations about people native to northern China. He said that the difference could, perhaps, be attributed to the "equanimity of the Chinese," but suggested that diet was probably a more important explanation. These observations led the American scientist Ancel Keys to wonder about the geographical variation in the incidence of CAD (Keys 1983). He initiated a massive study, published as *Seven Countries: A Multivariate Analysis of Death and Coronary Heart Disease*. This detailed work was begun in 1947 and eventually resulted in the study of a total of 12,763 men over 10 years in Yugoslavia, Finland, Italy, the Netherlands, Greece, the United States, and Japan. Conclusions are complex (on a scale with the study itself), but they show a clear relationship between the dietary intake of saturated fat and cholesterol and the incidence of coronary heart disease. Other between-population studies have shown that in countries where CVD rates are low, the serum LDL levels also tend to be low (for example, rural Japan and China compared to the United States and Finland).

Other investigations (following the lead of De Langen) have examined changes in CAD incidence with migration. Most have shown that when people move from a region with a low incidence of CAD to a region with a high incidence of CAD, and adopt the lifestyle of their new country, their likelihood of having CAD approaches that of the region to which they

have moved. For instance, Japanese people living in their native Japan, in Hawaii, and in San Francisco, have an increasingly higher incidence of CVD: Age-adjusted rates among Japanese people were 1.6 per 1,000 person-years in Japan, 3.0 in Hawaii, and 3.7 in San Francisco (Kato et al. 1973). Saturated fat intake as a percentage of calories in the three populations was 7, 23, and 26 percent, respectively. The dietary trends among these populations emphasize the importance of the type of fat intake: Foods high in saturated fat also tend to be high in cholesterol. What happened to Japanese people moving from Japan to Hawaii to San Francisco seems clear enough. As they adopted a progressively more Western diet with a higher intake of saturated fats, their incidence of heart disease increased.

Cholesterol-Lowering Intervention

If the changing of diet results in a changing rate of heart disease, then public education efforts to heighten personal awareness of cholesterol levels may be serving as a broad and effective means of intervention. There are some encouraging signs. From 1980 to 1987, in the Minneapolis–St. Paul area, there was a significant decrease in total serum cholesterol and in the numbers of individuals with significant hypercholesterolemia (greater than 5.04 micromoles [mmol] per liter or 195 mg per deciliter) (Burke et al. 1991). However, despite the downward trend of cholesterol in the study population as a whole, many individuals (67 percent) with cholesterol levels high enough to require dietary and/or medical therapy remained unaware of their condition.

On a national level, the National Health and Nutrition Examination Survey (NHANES), undertaken between 1976 and 1980 (NHANES II) and from 1988 to 1991 (NHANES III), found an increase in the number of people with total cholesterol values less than 5.17 mmol per liter (200 mg per deciliter) and a decrease in those with total cholesterol values greater than 6.21 mmol per liter (240 mg per deciliter).

Although these population-based observations may reflect national public-health efforts to reduce total and saturated fat and cholesterol, such hopeful trends probably do not completely explain the decline in mortality from CVD. During the 1980s in the United States, there was also a decline in the use of tobacco, a major risk factor for CAD. In addition, there were changes in the diagnosis and management of CAD, including more widely available cardiac catheterization, angioplasty, functional cardiac studies (for example, dipyridamole and thallium stress echocardiograms), and development and utilization of new medications (for instance, thrombolytics, cardio-specific beta-blockers, and angiotensin converting enzyme inhibitors). Thus, at present, the relative importance of population-based cholesterol lowering in bringing about a declining CVD death rate remains unknown.

Nonetheless, although many factors could alter mortality from CAD, the lowering of cholesterol would intuitively seem both important and appropriate. Prevention of CAD among a community of individuals can be viewed from two perspectives: primary prevention and secondary prevention.

Primary prevention. Efforts to prevent persons from developing a particular disease fall under the rubric of primary prevention. In the case of CAD, primary prevention implies reducing the disease incidence for people with risk factors but who have not yet developed it. The Lipid Research Clinics Program (1984) and the Helsinki Heart Study (Frick et al. 1987) were large, randomized, controlled clinical trials that indicated that reducing serum cholesterol in persons without known CAD reduces the incidence of CAD onset. However, these trials were short (less than 10 years) and did not (or could not) demonstrate that primary prevention lengthens the life span by delaying or preventing the onset of CAD. Individuals who have hypercholesterolemia as their only risk factor for CAD likely do not add many years to their life by reducing their cholesterol level (Browner, Westenhofe, and Tice 1991). Indeed, some studies have even suggested a rise in the incidence of violent deaths among individuals with the lowest cholesterol levels, although the relevance of such observations remains controversial.

Secondary prevention. On average, however, 3 to 4 years of life can be gained by individuals who have CAD and reduce their serum cholesterol by dietary and/or by medical means. This approach – intervention in the presence of known disease – defines secondary prevention.

The idea that diet has some relationship with heart disease, and that changing the diet can help people with heart disease, has been around for some time. What entered and left the body were long held as being critically important to its general functioning. Thus, it should come as no surprise to see comments about treating heart disease with diet from, essentially, the very first speculations about heart diseases in general (Fothergill 1776, cited in Leibowitz 1970). We should not attempt, however, to read these comments as having the same sort of specificity we now associate with statements about disease causation and diet.

For the early twentieth century, some insight into generally accepted notions of diet and heart disease may be derived by consulting a series of articles published in 1913 by the American Medical Association (AMA), the “cordial reception” of which led to their presentation in book form. The opinions expressed in these articles may be regarded as reflecting a general consensus, at least among allopathic physicians. Some of what was advised for the patient with heart disease seems logical (if not necessarily appropriate) to the

late-twentieth-century reader, such as the suggestion that coffee should be avoided or that obese patients should lose weight in order to help the functions of their hearts.

Other approaches, such as administering hydrochloric acid to aid in digestion, now seem clearly out of place. But what is most striking is the emphasis on the idea that diet should be individualized, at least in part because it is limited by “what the patient will do.” Rather than being dogmatic about what constitutes a “correct” diet, advisers would do better to tailor their recommendations to the specific social and cultural setting of the person under treatment.

Across the Atlantic, the noted British physician Sir Clifford Allbutt, Regius Professor at Cambridge University in England, was also concerned about individual variation in dietary suggestions (Allbutt 1915, particularly 238–54 of Vol. 1). He considered the question of overfeeding to be a relative one: What was a good diet for one person might be gluttony for another. He noted that cholesterol was associated with atherosclerosis, but he saw also the wide variation in the diets of people who developed atherosclerosis.

More recently, many trials have examined the value of cholesterol reduction in the management of CAD. These studies usually randomized patients with known CAD to receive some type of treatment (diet, drug, lifestyle change) versus “usual care” or placebo. The outcomes of interest in these studies tended to be occurrence of angina, myocardial infarction, death, and regression of atherosclerosis. Most subjects had total serum cholesterol concentrations greater than 5.00 mmol per liter. To determine regression or progression of atherosclerosis, baseline coronary angiograms were compared with those at the end of the study period. The angiograms were assessed with quantitative angiography and/or global score assessment. Both methods asked the same question: Are the areas of stenosis the same, less, or greater than at baseline, and is there a relationship between the extent of plaque regression and the amount of cholesterol lowering?

The Cholesterol-Lowering Atherosclerosis Study (CLAS) (Blankenhorn et al. 1987) investigated the effects of drug therapy in a randomized, placebo-controlled trial using men ages 40 to 59 years who had undergone coronary artery bypass graft surgery. After two years of treatment, patients who had been given the cholesterol-lowering drugs colestipol and niacin instead of placebos exhibited a reduction in total cholesterol, a reduction in LDL-cholesterol, and a significant amount of atherosclerotic plaque regression and preservation of native coronary arteries. However, the incidence of coronary events was no different between the two groups.

A fairly recent study by D. Ornish and colleagues (1990) explored the influence of lifestyle changes on reversal of CAD in a randomized controlled trial that followed men and women for one year. The treatment

group experienced aggressive lifestyle interventions, including the imposition of a very low cholesterol, strict vegetarian diet. Their overall dietary fat was reduced to 10 percent of calories, far lower than in the usual American diet, which contains 35 to 40 percent of calories from fat. The treatment group demonstrated regression of atherosclerotic lesions and reduction of serum cholesterol and LDL-cholesterol.

The investigation has been criticized for flaws in randomization. Moreover, the practicality of implementing such a drastic dietary intervention for the general population with CAD has also been questioned. However, the findings are provocative and suggest that a radical reduction in dietary fat can greatly influence serum cholesterol and plaque regression. The study subjects also received psychological and behavioral interventions and increased their exercise. Whether such changes can alter morbidity and mortality from CAD over one's lifetime to a greater extent than more conservative, yet aggressive, approaches to cholesterol lowering remains to be elucidated.

The St. Thomas Atherosclerosis Regression Study (STARS) (Watts et al. 1992) focused on 90 men with CAD and mild hypercholesterolemia for approximately three years. It investigated the effects of usual care versus two interventions: a low-cholesterol diet and a low-cholesterol diet with the cholesterol-lowering drug, cholestyramine. Both interventions reduced the frequency of cardiovascular events and the progression of coronary artery narrowing.

The Monitored Atherosclerosis Regression Study (MARS) (Blankenhorn et al. 1993) was a randomized, double-blind, placebo-controlled study of 270 men and women with CAD that evaluated the effect of a low-cholesterol, reduced-fat diet with and without the HMG-CoA reductase inhibitor lovastatin. This trial demonstrated regression of atherosclerotic plaques; however, there was no difference in cardiac events between the treatment and placebo groups. At first glance, these results may seem to contradict the STARS trial, but the diet in the control group of the STARS study was not one designed to lower plasma lipids. In the STARS trial, both the lipid-lowering diet and the diet plus cholestyramine resin reduced coronary events to the same degree. They were significantly different from the usual care group, but not from each other.

The data on the role of cholesterol, particularly LDL-cholesterol, in the pathogenesis of CAD is compelling. Epidemiological population studies strongly suggest a positive relationship between elevated serum cholesterol and mortality rates from CAD. Primary and secondary prevention trials support the contention that reducing LDL-cholesterol, through a lipid-lowering diet with or without adjuvant medication, alters disease progression in middle-aged men. It is less clear if therapy will significantly reduce the incidence of subsequent cardiac events. The trials have not demonstrated an impact on longevity. How-

ever, it is certainly possible that future studies will do so. Studies in the Scandinavian Simvastatin Survival Study (known as 4S) seem to show an overall decrease in the death rate for people who already had heart disease when treated with potent cholesterol-lowering drugs (Scandinavian Simvastatin Survival Study Group 1994).

Dietary Fat

The influence of diet on CAD may reflect the amount of fat a person ingests. But the type of fat consumed may be just as important as the amount. Cholesterol is a type of fat found only in animal cells, not in plant cells. Low-cholesterol diets stress not only a reduction in dietary cholesterol but also a reduction in overall dietary fat, especially saturated fat. This approach is necessary because animal fat is laden with cholesterol. Animal fat is also primarily a saturated fat. Thus, the current recommendations by several national advisory panels are to decrease the consumption of saturated, cholesterol-rich animal fat, and to increase the amount of monounsaturated and polyunsaturated, or plant-derived, fat. Because poultry and fish contain more polyunsaturated fat than beef or pork, the recommendations also suggest consuming more poultry and fish as protein sources.

In dietary recommendations made in 1970 in the "Report of the Inter-Society Commission for Heart Disease Resources" (which were incorporated into "Dietary Goals for the United States by the United States Senate Select Committee on Nutrition and Human Needs"), the consumption of fat was expected to decrease from 42 to 30 percent of calories. Since 1909, Americans had consumed an average of 600 mg of cholesterol per day, but with these recommendations, cholesterol consumption was projected to decrease to 300 mg per capita per day. Such recommendations were considered radical at the time, and those who proposed diets based on them were concerned about the ability of patients to consume such a diet while maintaining normal protein nutrition.

More recently, however, the National Research Council's 1992 recommendations have emphasized that in order to better meet the national goal of reducing total cholesterol concentrations to less than 5.04 mmol per liter, daily calories from fat should not exceed 30 percent of total calories, and two-thirds of that fat should be monounsaturated and polyunsaturated.

Saturated versus Unsaturated Fats

Can the rise in CVD mortality seen in countries like the United States and Scandinavia be attributed to the type of dietary fat, as well as to dietary cholesterol? The answer to the question seems to be yes. In the first half of this century, although U.S. CVD mortality rates rose dramatically, the increase in consumption

of total and saturated fat was much more modest. However, during the same period, consumption of polyunsaturated fat rose two to three times (Friend 1967; Page and Marston 1979). In Europe during World War II, as wartime deprivation curtailed animal fat and, in fact, total fat consumption, the number of deaths due to CAD fell dramatically. In the 1970s, rural Romanians consumed 900 mg of cholesterol per day – 50 percent more than Americans at the time – but Romanian CVD mortality rates were approximately 20 percent lower, perhaps because their dietary fat intake was 30 percent less than the average in the United States (WHO 1976).

Such observations implicate dietary fat, not just dietary cholesterol, in atherogenesis. Further, the nature of the fat (that is, saturated or unsaturated) may be important. Whether a fat is saturated or not has to do with how many hydrogen atoms are bound to carbon. Carbon atoms form the “backbone” of a fat molecule and can maximally bind to 4 other atoms. If 4 binding sites are used, the carbon atom is said to be “saturated.” If only 2 binding sites are used, the carbon atom forms a double bond with another carbon atom and is said to be “unsaturated.” Saturated fatty acids have no carbon double bonds, monounsaturated fatty acids have one carbon double bond, and polyunsaturated fatty acids have two or more carbon double bonds.

Epidemiological studies have observed that mortality from CVD is lower in southern European countries. There, total fat consumption is not remarkably low compared to other industrialized nations, but there is a greater consumption of monounsaturated fats, such as olive oil.

In 1928, two Chinese biochemists suggested that the low rate of CAD in China might be explained by the polyunsaturated fat linoleic acid (Snapper 1963). Others have later shown that this substance can lower serum cholesterol (Ahrens et al. 1959). In the United States, increased consumption of linoleic acid, primarily through greater consumption of corn oil products, parallels the decline in population serum cholesterol concentrations.

Marine Oils

Fish oils appear to lower serum lipids in animals and humans (Bronte-Stewart et al. 1956; Nelson 1972). In the 1970s, epidemiologists attempted to explain a striking difference in the incidence of CAD between Eskimos, who rarely suffer from CVD, and Danes. They found a major dietary difference, not in total fat consumption but in the amount of fish oil consumption. Eskimos who have diets high in omega-3 fatty acids (found in fish oil) have prolonged bleeding times, as well as a decreased number and aggregation of platelets – all features associated with a reduced incidence of coronary thrombosis. Several hypotheses have been presented to explain the antiatherogenic effects of marine oils. They include altered plasma

cholesterol and triglyceride concentrations, altered metabolism of prostaglandins and leukotrienes, and a variety of other physiological responses (Zhu and Parmley 1990).

Laboratory experiments have suggested possible mechanisms for the effects of marine oils. Animals fed a diet high in omega-3 long-chain polyunsaturated fatty acids and dietary cholesterol had a reduction in the number and size of atherosclerotic lesions. This may have resulted from chemical changes that affect how blood cells adhere to lesion-prone areas. Associated with these physiological changes were alterations in the prostaglandin synthetic pathway reflected by decreased thromboxane A₂, increased prostacyclin, decreased leukotriene B₄, and increased leukotriene B₅. Most of these studies have also demonstrated significant lowering of plasma total cholesterol, LDL-cholesterol, and triglyceride levels. HDL-cholesterol has generally been unchanged.

Clinical investigations have shown that the consumption of fish oil in normal volunteers and in patients with hyperlipidemia can remarkably decrease plasma triglyceride levels, with inconsistent effects on plasma cholesterol and HDL-cholesterol. Perhaps as a result, fish oil supplements were widely marketed in the 1980s. But although epidemiological studies of Eskimos suggest a chemoprotective effect of fish oil, it is unclear that increasing fish oil in an American or northern European diet will impact CVD mortality rates. In addition, problems with fish oil supplements, including a bad flavor and unpalatable belching, pose compliance problems for patients with CVD risk.

Other Dietary Components

Dietary Protein

Sources of dietary protein include meat (for example, beef, pork, lamb, and wild game), poultry (such as chicken, turkey, and quail), fish, seafood, dairy products, eggs, grains, legumes (for example, peanuts, soy, and dried beans), nuts, and seeds. Depending on one's culture, religion, ethnicity, and other socioeconomic variables, the primary daily protein source will vary. Different protein sources contain varying amounts and types of fat. The discussion in the section “Dietary Fat” has pointed out that saturated animal fat is most relevant to hyperlipidemia, unlike plant lipids, which are unsaturated.

Milk protein has been implicated as an important nutrient in atherogenesis for more than 20 years, most notably in S. Seely's epidemiological analyses (1981, 1988). By reviewing the CVD mortality rates in men from 24 countries along with food consumption data, he found a significant correlation between CVD and the consumption of unfermented milk proteins, the only exception being cheese (1981). In 1988, Seely studied food consumption patterns of men and women in 21 countries that might be associated with

CVD mortality. He observed significant positive correlations between CVD mortality and milk, milk products, sugar, and oats. There were negative correlations with fish proteins, vegetable proteins, and fish fat.

Proteins are made of amino acids, and one of those amino acids - arginine - has become the subject of intense scrutiny. The discovery in the 1980s and 1990s that mammals endogenously synthesize nitric oxide, a toxic substance, has led to an enormous body of literature describing the physiology, immunology, and biochemistry of this substance. It is produced by the enzymatic conversion of L-arginine to nitric oxide by nitric oxide synthetase. This reaction occurs in several tissue types, including endothelium, macrophages, and the brain. Nitric oxide produced in endothelium causes relaxation of the vessel and, hence, vasodilation.

L. Castillo and colleagues (1995) demonstrated that up to 16 percent of dietary arginine may be converted to nitric oxide in healthy volunteers. Several investigators have reported abnormal coronary arterial vasodilation, probably related to nitric oxide production in patients with increased cholesterol (Creaiger et al. 1990), hypertension (Linder et al. 1990), and CAD (Zeiber et al. 1991; Egashira et al. 1993). M. Jeserich and colleagues (1992) found significantly lower plasma L-arginine concentrations in patients with hypercholesterolemia (greater than 270 mg per deciliter), as compared to patients with normal cholesterol levels (less than 220 mg per deciliter). The relationship of these findings to CVD pathogenesis remains to be elucidated. It is unclear whether this difference in arginine levels is because of diet or because of the process of atherogenesis.

Dietary Carbohydrates and Fiber

Carbohydrate foods provide the bulk of daily energy needs. Animal studies have found that if the ratio of energy to protein is low, serum cholesterol tends to be lower. In humans, decreased energy intake may be associated with less obesity, a lower incidence of diabetes, and reduced cholesterol (total and LDL) (Kannel 1987).

Thus, modification of these atherogenic risk factors by modifying energy intake may be beneficial. The other main dietary energy source is fat, which provides about twice the number of calories per gram as carbohydrate does. Thus, to reduce overall energy intake and reduce atherosclerotic risk, it is desirable to obtain energy calories primarily from carbohydrate sources.

The association of fiber with hyperlipidemia has also received attention. Soluble fiber consumption is associated with reduced serum cholesterol. D.J.A. Jenkins and colleagues (1993) studied soluble fiber supplementation in conjunction with a lipid-lowering diet in 43 healthy men and women of normal weight. They observed a significant reduction in total cholesterol in both groups compared to base-

line, but the soluble fiber group attained an even greater reduction.

Vitamins

"Antioxidant" nutrients have recently gotten a great deal of media and scientific coverage. These compounds include beta-carotene, vitamin E, and vitamin C. There are several reasons for this interest. LDL-cholesterol is oxidatively modified through chemical mediators released from the macrophage during atherosclerotic pathogenesis. Nonnutrient antioxidant substances (for example, probucol, butylated hydroxytoluene) inhibit the progression of atherogenesis in rabbits, presumably due to an alteration in oxidatively modified LDL-cholesterol (Bjorkhem et al. 1991; Mao et al. 1991). Antioxidants such as vitamin E work by "grabbing" a hydrogen atom. Having done so, they are now the oxidized molecule, and they have thus prevented another moiety, such as LDL, from becoming oxidized. Epidemiological studies suggest that people with high plasma levels of vitamin E have a lower risk of CVD. Similar results have been obtained when examining beta-carotene intake and CVD incidence. In the Nurses' Health Study, CVD risk was reduced 30 to 40 percent in those individuals with high calculated intakes of vitamin E or beta-carotene. The data on vitamin C were unconvincing (Stampfer et al. 1987; Stampfer et al. 1993).

M. Abbey, P. J. Nestel, and P. A. Baghurst (1993) investigated LDL oxidation in nonsmoking men and women who were randomized to receive either a placebo or an "antioxidant" vitamin supplement (18 mg beta-carotene, 250 mg d-alpha-tocopheryl succinate, and 12 mg zinc) for 6 months. During the 6-month study period, the supplemented group's LDL was less able to oxidize than the control group. I. Jialal and S. M. Grundy (1993) also demonstrated a decrease in LDL oxidation rate in men supplemented with vitamin E.

There are, however, several problems with recommending supplements of vitamin E or any other of the "antioxidants." First, many of these vitamins have biochemically important interactions. For instance, vitamin C may reduce vitamin E needs by reducing oxidized vitamin E. The chemically reduced vitamin E can then be "recycled" and used again as an antioxidant. Second, no well-controlled randomized clinical trials have determined what dose of antioxidant may be needed to accomplish atherosclerosis prevention without adverse side effects. We do not even know which nutrient (or mixture of nutrients) should be tested. Further, it is not known if *in vitro* LDL oxidation is the best marker for what these nutrients may do in atherogenesis.

Thus, while it is probable that some of these nutrients play important roles in lipid metabolism (which may affect CVD), we simply do not know enough about them to make sweeping recommendations. If individuals would reduce their dietary fat to 30 per-

cent of calories and increase their consumption of complex carbohydrates to 55 percent of calories (including fruits and vegetables), they would likely increase their ingestion of these “antioxidants,” with numerous overall health benefits and without risk.

Alcohol

Several studies suggest a possible benefit of alcohol ingestion. LDL-cholesterol levels appear to be lower in individuals who consume moderate amounts of alcohol: 1 to 3 bottles of beer, glasses of wine, or shots of liquor per day. At higher levels of alcohol ingestion, LDL is further reduced, probably due to the replacement of alcohol for energy calories at the expense of ingesting foods containing cholesterol and fat. Alcohol is the only substance humans ingest (nonpharmacologically) that can raise HDL-cholesterol, a finding that probably explains the beneficial effect of alcohol ingestion in reducing CVD mortality. However, in addition to its beneficial effects on blood lipids, alcohol is a drug with significant potential for abuse. Thus, most public health officials, physicians, and scientists are concerned about advising alcohol as a therapeutic modality for hyperlipidemia and CVD prevention.

Miscellaneous Dietary Components

S. Warshafsky, R. S. Kamer, and S. L. Sivak (1993) examined five placebo-controlled randomized trials on the effects of garlic supplementation on serum cholesterol. About one-half to one clove of garlic per day decreases serum cholesterol by about 9 percent. Proposed mechanisms of garlic's effects on blood lipids include increased bile acid excretion and reduced HMG-CoA reductase activity in the liver. However, carefully controlled studies of garlic remain to be done.

In concert with the antioxidant hypothesis, certain minerals have been implicated in atherogenesis. Selenium is able to act as a free-radical scavenger through its role as a cofactor in the enzyme glutathione peroxidase. However, the epidemiological studies on selenium consumption and CVD mortality do not bear out a probable relationship.

Conclusion

We have reviewed several dietary components that have been implicated in the pathogenesis of cardiovascular disease. The contribution of most nutrients appears to revolve around their ability to influence the serum lipid profile – total cholesterol, LDL-cholesterol, HDL-cholesterol, triglycerides, and the various apoproteins.

There is a need to look beyond serum lipids, however. Besides influencing a specific set of measurable laboratory parameters, nutrients must impact the pathogenesis of disease through biochemical and molecular alterations at the cellular level. This milieu of biochemical reactions and interactions is ultimately

the most consequential to the dietary pathogenesis, prevention, and therapy of cardiovascular disease. We will doubtless continue to receive information about the important benefits of “nutrient A” on “function B,” and physicians, scientists, patients, and the public will try to make sense of these assertions.

Yet one must remember that nutrients are but small components of the whole chemical “package” that makes up a particular food or group of foods. Other properties of additional components in food may work synergistically or antagonistically, or act as innocent bystanders in cardiovascular disease chemoprotection and pathogenesis. It is unlikely that in years to come a single nutrient – will be uncovered as “the cause” of CAD. More likely, further investigation will describe the subtle interactions of various dietary components with an individual's genome, in concert with other cardiovascular risk factors, such as tobacco use, diabetes mellitus, hypertension, hypoestrogenemia, stress, and obesity. Researchers probably will demonstrate that it is not just one or two nutrients or dietary factors that are critical, but the diet as a whole.

Consider an 88-year-old, generally healthy man who had a verified consumption of 20 to 30 whole eggs per day (Kern 1991). Despite this incredibly high daily cholesterol intake (12,953 micromoles), his total cholesterol was only 5.18 mmol per liter (200 mg per deciliter), and his LDL-cholesterol was 3.68 mmol per liter (142 mg per deciliter). It appears that he had compensated for his dietary excess by excreting excess cholesterol bound in bile acids in his stool. Thus, how an individual utilizes the diet biochemically may be as important as the diet itself.

We are not yet in a position to predict which people will do well with high cholesterol intake and which will not. In the meantime, it is prudent to follow the recent dietary guidelines of the National Research Council, which stress a diet containing 15 to 20 percent of calories from protein, 25 to 30 percent of calories from dietary fat, and the remaining calories from complex carbohydrates, such as whole grains, vegetables, and fruits.

These guidelines were put forward as death rates from CAD were falling, and there is no shortage of advocacy groups ready and willing to take credit for this reduction. But it is most probably the result of many lifestyle changes, among them declining use of tobacco (at least in the United States, when considered for all members of the population combined); better emergency medical services; improved in-hospital treatments; and, of course, dietary modifications. As noted, change in diet has a particular appeal as an explanation because it is the only one of these factors that daily affects each and every member of society. And, as the new millennium dawns, it would seem that dietary intervention to prevent CAD has proved to be a successful means of secondary prevention for those people who have already incurred a

cardiac event, even if the utility of such intervention as a means of primary prevention, applied to entire populations, remains controversial.

We should also bear in mind that the debate over diet and heart disease is being conducted in a very public arena, and an arena in which diet and heart disease are only one aspect of a more general tension over questions focusing on personal risk and responsibility. In the case of heart disease, these questions have stirred a broad-based popular reaction. No longer are discussions of diet and the heart confined to the pages of medical journals, but they can now be found in practically every issue of widely circulating newspapers and magazines. One good example is an extensive and detailed analysis questioning the overall importance of cholesterol in heart disease – indeed, dismissing “The Cholesterol Myth” – in the *Atlantic Magazine* (Moore 1989a). In the same year that this essay appeared, it was also published in expanded form as a book (Moore 1989b).

Public skepticism about the importance of diet in heart disease is also apparent in cartoons. One, in the *New Yorker* magazine (January 16, 1989: 39), shows a man starting to eat a huge oatmeal muffin. The caption reads: “Wellness update: Thirty-year-old man starting on the twenty-five-thousand-pound oat-bran muffin he must consume over forty years in order to reduce significantly his risk of death from high cholesterol.”

This and countless other less-than-reverent cartoons constitute widespread cultural markers that express a number of popular reactions to the diet-heart hypothesis. First, they ventilate frustration at the gospel of eating for one’s heart (or for health in general). And given a culture in which the average person is daily bombarded with images of food, with much of the easiest food to acquire and consume that which is likely to be least beneficial for preventing heart disease, this frustration is understandable. Such cartoons also reflect a general lack of conviction that altered diets will “work” against heart disease – a skepticism that primary prevention trials have thus far failed to address convincingly. At a lay level, most people know (or have heard of) someone who has lived to a ripe old age while engaging in near-constant dietary indiscretions, whereas someone who ate “right” and stayed fit may have suffered an early death. Epidemiologists might discount the relevance of such anecdotes, but they can have a major impact on popular perceptions of risk and disease, especially when the behavior that is being advocated may be neither easy nor (seemingly) pleasurable.

The popular reaction against dietary constraints also raises fundamental questions of personal versus public responsibility. If someone has CAD, whose fault is it? The choices people make about what to eat are limited by the cultural world in which they live, so to what extent is the larger society to be held responsible? And, if one is personally accountable for what foods are consumed, then should a “healthy lifestyle”

make a difference in terms of how much one pays for health or life insurance?

Finally, what is the disease here? Is it CAD, or is it high cholesterol? We should not lose sight of the fact that blood lipids are merely surrogates for what is most important – the disability and death that come from CAD and CVD. Keeping track of these differences is important. First, it can help to maintain focus on the ultimate goals of therapy and not allow us to become sidetracked by that which is easier to effect. Second, it will prevent the labeling of a large percentage of the population as “diseased” simply because of a high lipid profile. The answers to the questions raised in this chapter are unlikely to come purely from the accumulation of more and more data, because the prevention controversies are also fueled by “hidden ideological, structural, and professional factors” (Aronowitz 1994).

Melissa H. Olken
Joel D. Howell

Bibliography

- Abbey, M., P. J. Nestel, and P. A. Baghurst. 1993. Antioxidant vitamins and low-density lipoprotein oxidation. *American Journal of Clinical Nutrition* 58: 525–32.
- Ahrens, E. H., W. Insull, J. Hirsch, et al. 1959. The effect on human serum-lipids of a dietary fat, highly unsaturated but poor in essential fatty acids. *Lancet* 1: 117–19.
- Allbutt, Clifford. 1915. *Diseases of the arteries, including angina pectoris*. 2 vols. London.
- Armstrong, B. K., J. I. Mann, A. M. Adelstein, and F. Eskin. 1975. Commodity consumption and ischemic heart disease mortality, with special reference to dietary practices. *Journal of Chronic Disease* 28: 455–67.
- Aronowitz, Robert. 1994. Trouble in prevention. *Journal of General Internal Medicine* 9: 475–8.
- Bjorkhem, I., A. Henriksson-Freyschuss, O. Breuer, et al. 1991. The antioxidant butylated hydroxytoluene protects against atherosclerosis. *Atherosclerosis and Thrombosis* 11: 15–22.
- Blankenhorn, D. H., S. P. Azen, D. M. Kramsch, et al. 1993. Coronary angiographic changes with lovastatin therapy. *Annals of Internal Medicine* 119: 969–76.
- Blankenhorn, D. H., S. A. Nessim, R. L. Johnson, et al. 1987. Beneficial effects of combined colestipol-niacin therapy on coronary atherosclerosis and coronary venous bypass grafts. *Journal of the American Medical Association* 257: 3233–40.
- Bronte-Stewart, B., A. Antonis, L. Eales, and J. F. Brock. 1956. Effects of feeding different fats on serum-cholesterol level. *Lancet* 2: 521–6.
- Browner, Warren S., Janice Westenhause, and Jeffrey A. Tice. 1991. What if Americans ate less fat? A quantitative estimate of the effect on mortality. *Journal of the American Medical Association* 265: 3285–91.
- Burke, G. L., J. M. Sprafka, A. R. Folsom, et al. 1991. Trends in serum cholesterol levels from 1980–1987. *New England Journal of Medicine* 324: 941–6.
- Castillo, L., M. Sanchez, J. Vogt, et al. 1995. Plasma arginine, citrulline, and ornithine kinetics in adults, with observa-

- tions on nitric oxide synthesis. *American Journal of Physiology* 268: E360-7.
- Creager, M. A., J. P. Cooke, M. E. Mendelsohn, et al. 1990. Impaired vasodilation of forearm resistance vessels in hypercholesterolemic humans. *Journal of Clinical Investigation* 86: 228-34.
- Dalderup, L. M. 1973. Ozone, vitamin E, fatty acids, prostaglandins, atherosclerosis and its complications (letter to the editor). *Archives of Environmental Health* 27: 58.
- Dawber, D. V., and L. N. Katz. 1943. Experimental atherosclerosis in the chick. *Archives of Pathology* 36: 473-92.
- Egashira, K., T. Inou, Y. Hirooka, et al. 1993. Evidence of impaired endothelium-dependent coronary vasodilation in patients with angina pectoris and normal coronary angiograms. *New England Journal of Medicine* 328: 1659-64.
- Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults. 1993. Summary of the second report of the National Cholesterol Education Program (NCEP). *Journal of the American Medical Association* 269: 3015-23.
- Frick, M. H., O. Elo, K. Hapka, et al. 1987. Helsinki heart study: Primary-prevention trial with gemfibrozil in middle-aged men with dyslipidemia: Safety of treatment, changes in risk factors, and incidence of coronary heart disease. *New England Journal of Medicine* 817: 1237-45.
- Friend, B. 1967. Nutrients in the United States food supply: A review of trends, 1909-1913 to 1965. *American Journal of Clinical Nutrition* 20: 907-14.
- Godfried, S. L., G. F. Combs, J. M. Saroka, and L. A. Dillingham. 1989. Potentiation of atherosclerotic lesions in rabbits by a high dietary level of vitamin E. *British Journal of Nutrition* 61: 607-17.
- Jenkins, D. J. A., T. M. S. Wolever, A. V. Rao, et al. 1993. Effect on blood lipids of very high intakes of fiber in diets low in saturated fat and cholesterol. *New England Journal of Medicine* 329: 21-6.
- Jeserich, M., T. Munzel, H. Just, and H. Drexler. 1992. Reduced plasma L-arginine in hypercholesterolaemia (letter). *Lancet* 339: 561.
- Jialal, I., and S. M. Grundy. 1993. Effect of combined supplementation with alpha-tocopherol, ascorbate, and beta carotene on low-density lipoprotein oxidation. *Circulation* 88: 2780-6.
- Kannel, W. B. 1987. New perspectives on cardiovascular risk factors. *American Heart Journal* 114: 213-19.
- Kannel, W. B., W. P. Castelli, T. Gordon, and P. M. McNamara. 1971. Serum cholesterol, lipoproteins, and the risk of coronary heart disease: The Framingham Study. *Annals of Internal Medicine* 74: 1-12.
- Kato, H., J. Tillotson, M. Z. Nichamen, et al. 1973. Epidemiologic studies of coronary heart disease and stroke in Japanese men living in Japan, Hawaii and California: Serum lipids and diet. *American Journal of Epidemiology* 97: 372-85.
- Kern, F. 1991. Normal plasma cholesterol in an 88-year-old man who eats 25 eggs a day. *New England Journal of Medicine* 324: 896-9.
- Keys, Ancel. 1980. *Seven countries: A multivariate analysis of death and coronary heart disease*. Cambridge, Mass.
1983. From Naples to seven countries – A sentimental journey. *Progress in Biochemical Pharmacology* 19: 1-30.
- Keys, Ancel, H. L. Taylor, H. Blackburn, et al. 1963. Coronary heart disease among Minnesota business and professional men followed for fifteen years. *Circulation* 28: 381-95.
- Kummerow, F. A. 1979. Nutrition imbalance and angiotoxins as dietary risk factors in coronary heart disease. *American Journal of Clinical Nutrition* 32: 58-83.
- Leary, Timothy. 1935. Pathology of coronary sclerosis. *American Heart Journal* 10: 328-37.
- Leibowitz, J. O. 1970. *The history of coronary heart disease*. Berkeley, Calif.
- Linder, L., W. Kiowski, F. R. Buhler, and T. P. Luscher. 1990. Indirect evidence for release of endothelium-derived relaxing factor in human forearm circulation in vivo: Blunted response in essential hypertension. *Circulation* 81: 1762-7.
- Lipid Research Clinics Program. 1984. The lipid research clinics coronary primary prevention trial results: I. Reduction in incidence of coronary heart disease. *Journal of the American Medical Association* 251: 351-64.
- Mao, S. J., M. T. Yates, A. E. Rechten, et al. 1991. Antioxidant activity of probucol and its analogues in hypercholesterolemic Watanabe rabbits. *Journal of Medicinal Chemistry* 34: 298-302.
- Moore, Thomas J. 1989a. The cholesterol myth. *Atlantic Magazine* 264 (September): 37-40.
- 1989b. *Heart failure: A critical inquiry into American medicine and the revolution in heart care*. New York.
- Nelson, A. M. 1972. Diet therapy in coronary disease: Effect on mortality of high-protein, high-seafood, fat-controlled diet. *Geriatrics* 24: 103-16.
- NIH Consensus Development Panel. 1993. Triglyceride, high-density lipoprotein, and coronary heart disease. *Journal of the American Medical Association* 269: 505-10.
- Ornish, D., S. E. Brown, L. W. Scherwitz, et al. 1990. Can lifestyle changes reverse coronary heart disease? *Lancet* 336: 129-33.
- Osborne, Oliver T. 1913. *Disturbances of the heart: Discussion of the treatment of the heart in various disorders*. Chicago.
- Page, L., and R. M. Marston. 1979. Food consumption pattern – U.S. diet. In *Proceedings of the conference on the decline in coronary heart disease mortality*, ed. R. J. Havlik and M. Feinleib, 236-43. U.S. Dept. of Health, Education and Welfare, Public Health Service, National Institute of Health, Washington, D.C. (NIH Pub. No. 79-1610).
- Report of Inter-Society Commission for Heart Disease Resources. 1970. *Circulation* 42:A-55.
- Sabate, J., G. E. Fraser, K. Burke, et al. 1993. Effects of walnuts on serum lipid levels and blood pressure in normal men. *New England Journal of Medicine* 328: 603-7.
- Scandinavian Simvastatin Survival Study Group (4S). 1994. Randomized trial of cholesterol lowering in 4444 patients with coronary heart disease: The Scandinavian Simvastatin Survival Study. *Lancet* 344: 1383-9.
- Schwartz, C. J., A. J. Valente, and E. A. Sprague. 1993. A modern view of atherogenesis. *American Journal of Cardiology* 71: 9B-14B.
- Seely, S. 1981. Diet and coronary disease: A survey of mortality rates and food consumption statistics of 24 countries. *Medical Hypotheses* 7: 907-18.
1988. Diet and coronary arterial disease: A statistical study. *International Journal of Cardiology* 20: 183-92.
- Sempos, C. T., J. I. Cleeman, M. D. Carroll, et al. 1993. Prevalence of high blood cholesterol among US adults. *Journal of the American Medical Association* 269: 3009-14.
- Snapper, I. 1963. Diet and atherosclerosis: Truth and fiction. *American Journal of Cardiology* 11: 283-9.
- Stampfer, M. J., G. A. Colditz, W. C. Willett, et al. 1987. Coronary heart disease risk factors in women: The Nurses' Health Study experience. In *Coronary heart disease in*

- women, ed. E. Eaker, B. Packard, N. K. Wenger, et al., 112–16. New York.
- Stampfer, M. J., C. H. Hennekens, J. E. Manson, et al. 1993. Vitamin E consumption and risk of coronary heart disease in women. *New England Journal of Medicine* 328: 1444–9.
- Steinberg, D. 1991. Alcohol and atherosclerosis. *Annals of Internal Medicine* 114: 967–76.
1992. Antioxidants in the prevention of human atherosclerosis: Summary of the proceedings of a National Heart, Lung, and Blood Institute workshop: September 5–6, 1991, Bethesda, Maryland. *Circulation* 85: 2338–43.
- U.S. Senate Select Committee on Nutrition and Human Needs. 1977. *Dietary Goals for the United States*. Washington, D.C.
- Vogel, Julius. [1847] 1847. *The Pathological anatomy of the human body*, trans. G. E. Day. Philadelphia, Pa.
- Von Schacky, C. 1987. Prophylaxis of atherosclerosis with marine omega-3 fatty acids. *Annals of Internal Medicine* 107: 890–9.
- Walldius, G., J. Regnstrom, J. Nilsson, et al. 1993. The role of lipids and antioxidative factors for development of atherosclerosis. *American Journal of Cardiology* 71: 15B–9B.
- Warshafsky, S., R. S. Kamer, and S. L. Sivak. 1993. Effect of garlic on total serum cholesterol: A meta-analysis. *Annals of Internal Medicine* 119: 599–605.
- Watts, G. F., B. Lewis, J. N. H. Brunt, et al. 1992. Effects on coronary artery disease of lipid-lowering diet, or diet plus cholestyramine, in the St. Thomas's Atherosclerosis Regression Study (STARS). *Lancet* 339: 563–9.
- WHO (World Health Organization). 1976. *Myocardial infarction community registers*. Copenhagen.
- Zeiber, A. M., H. Drexler, H. Wollschlager, and H. Just. 1991. Modulation of coronary vasomotor in humans: Progressive endothelial dysfunction with different early stages of coronary atherosclerosis. *Circulation* 83: 391–401.
- Zhu, B-Q., and W. W. Parmley. 1990. Modification of experimental and clinical atherosclerosis by dietary fish oil. *American Heart Journal* 119: 168–78.

IV.F.4 🐉 The Cardiovascular System, Coronary Artery Disease, and Calcium: A Hypothesis

The Circulation

The central organ of the human circulatory system, the heart, must be among the most remarkable creations of nature. In the longest-living individuals, it works continuously for a hundred or more years, executing something like 4,000 million working strokes and moving 350,000 cubic meters of blood, enough to make a small lake. In individuals who die of heart disease, it is not, as a rule, the heart itself that fails but some auxiliary mechanism, like one of its arteries, or the pacemaker. If an artery, supplying a small part of the heart, is blocked, the tissues receiving oxygen and nutrients from that vessel die. If the

area involved is not so large as to endanger the entire heart, the damage is gradually repaired by the immune system. The dead cells are removed but cannot be replaced; the gap they leave is filled with scar tissue. While the repair is carried out, the heart continues to work.

Among other remarkable properties of the heart is a virtual immunity from cancer and a good resistance to inflammatory diseases. When the body is at rest, the heart contracts approximately once every second. Its contraction – the systole – lasts about one-third of a second; its relaxation period – the diastole – two-thirds of a second. During hard physical exercise, the heart rate increases about three times.

The arterial system is like a many-branched tree. Its trunk, the aorta, is about three centimeters in diameter at its origin. The branches become progressively smaller and end in a network of capillaries of microscopic size. On the return side, blood is collected by small venules, which join to form veins and end in two large venous trunks. The entire length of the system is more than enough to encircle the earth.

The function of the circulatory system is to deliver nutrients to the cell population of the body and to collect their waste products. The system is water-borne, but a very small quantity of water carries a comparatively enormous cargo. The volume of water in the blood, and that of the substances dissolved or floating in it, are approximately equal.

A water-based transport system can readily transport water-soluble substances, but those are only a small part of the nutrients carried in blood. The most difficult technical problem is the transport of gases, notably oxygen and carbon dioxide. These are soluble, but only sparsely so in water. If they were carried only in aqueous solution, a very large quantity would be required for their transport. The problem is solved by the use of iron porphyrin, a substance that can readily take up oxygen and equally readily release it. The porphyrins are incorporated in large hemoglobin molecules, with molecular weight of 64,000, and these, in turn, are incorporated into red blood cells. It is interesting to note that the mammalian type of hemoglobin is the end product of a long process of development in evolutionary history. Invertebrate blood contains a large variety of hemoglobin-like substances (erythrocruorins), with molecular weights ranging from 17,000 to 1,250,000, in some cases copper-based compounds being used instead of iron. Red blood cells constitute about 40 percent of the blood by volume. Allowing an equal quantity of water for their flotation, 80 percent of the total capacity of the circulatory system is engaged in the task of oxygen transport.

Similar difficulties are presented by the transport of lipids and other hydrophobic substances. If they were simply released in water, they would stick to each other and to vessel walls. There are two solutions to the problem. One is the conjugation of a

water-insoluble molecule with one or more other molecules to form a water-soluble complex. The other expedient is the use of carrier proteins. These are large molecules in comparison with the lipids they carry, serving, in effect, as packing cases for them. The protein molecule turns a hydrophilic outer surface to the watery medium in which it floats, and can pack a large quantity of hydrophobic molecules inside, shielded from contact with water.

The protein carrier with its lipid cargo constitutes a lipoprotein. As the carrier protein is heavier than water, the lipid it carries is lighter; the specific weight (or density) of their combination depends on the proportion of the two constituents. When the cargo of transport proteins is mainly cholesterol, they constitute low-density lipoproteins (LDL). Some lipoproteins carry cholesterol and other lipids from dead cells to the liver for reuse or excretion. They transport a small quantity of lipids; hence, they constitute high-density (HDL) lipoproteins.

The aqueous portion of the blood, the plasma, carries a variety of organic and inorganic substances, such as blood sugar (glucose) and salts. Among them are three kinds of proteins, albumin, globulin, and fibrinogen, serving various functions. The chemical messengers, hormones, which have to be transported from organs that produce them to receptors in other parts of the body, are always in transit in the blood, and so are enzymes, vitamins, and the like.

Moreover, the circulatory system is the highway that invading microorganisms try to use for spreading in the body, and cancers also spread by blood-borne fragments. To cope with these hazards, the system is defended by the immune mechanism, notably by several types of white blood cells (polymorphonuclear leukocytes, lymphocytes, and monocytes). The function of these cells is to seek out, ingest, and kill invading microorganisms, penetrate areas of infection, and phagocytose foreign matter and dead cells. The circulatory system is also the target organ of several plant toxins, some of which can act as vasodilators or constrictors, blood coagulants, or anticoagulants. The rye fungus ergot, for example, is such a powerful constrictor of peripheral blood vessels that, in the past, it created numerous epidemics in countries where rye was an important food item. The toxin ergotamine, mixed with rye flour, was capable of cutting off the circulation of the limbs, causing death by gangrene.

Apart from the effect of poisons upon them, arteries are subject to several other disorders (see the section "Specific Hazards of the Circulatory System"). Larger arteries are more vulnerable than small arteries. Veins are not entirely trouble free but, in comparison with arteries, are notable for their smooth functioning and resistance to disease. They do, of course, work under less onerous conditions than arteries.

Not surprisingly, arterial disorders are among the leading causes of mortality in advanced countries with older populations. In a few countries, they are

responsible for more mortality than all other causes combined. The current worldwide yearly toll of coronary artery disease is about two million deaths, to which cerebrovascular disease (strokes) add half a million. Disorders of the peripheral circulation can cause deaths, but since the elimination of ergotism, the numbers are small.

Specific Hazards of the Circulatory System

The foregoing review attempts to show the complexity and potential vulnerability of the circulation system. However, the exact causes of the most important arterial disorders are still not known, and the intent in this essay is to highlight a few specific hazards that could be relevant to their pathogenesis.

When the heart contracts, it compresses its own blood vessels, the coronary arteries, so that although it supplies all other organs in the body with blood, it cannot supply itself. The perfusion of the heart is effected by an auxiliary mechanism, working on the same principle as a flywheel in a mechanical system by drawing energy from a prime mover in one part of a power cycle and returning it in another part. The corresponding mechanism in the circulation is an elastic reservoir, consisting of the aorta and parts of its large-branch arteries. During systole the walls of the reservoir are distended, storing energy in their stretched elastic tissues; during diastole they contract, generating sufficient pressure to maintain blood flow in the circulation, including the by-then-relaxed coronary arteries. This arrangement, however, makes the nutrition of the heart dependent on the elasticity of the aorta and the other arteries that constitute the elastic reservoir. Several aspects of the aging process can bring about the deterioration of the elastic properties of tissues, representing a special hazard for the heart.

A similar problem arises in connection with the blood supply of the artery wall. The walls of arteries have a cell population that must be supplied with oxygen and nutrients like any other tissue. The difficulty in supplying them arises from the fact that the artery wall is compressed by the blood flowing through it, with a radial pressure gradient across it, the inside being more compressed than the outside. Large arteries have their own blood vessels, the *vasa vasorum*, which enter them on the outside and perfuse the outer half of the artery wall. They cannot perfuse the inner half, because the pressure available to force blood through these small arteries is insufficient to overcome the compression of that part of the artery. Consequently, the inner half of the artery wall must obtain oxygen and nutrients from blood flowing through the artery. The technical problems associated with this process are probably the basic causes of arterial disorders.

Oxygen and water-soluble nutrients can apparently diffuse through the layer of cells – the endothelium – that constitute the inner lining of the artery

wall, and there are no known disorders caused by deficiencies of such nutrients. The main difficulty is the transfer of lipids from arterial blood to tissues in the inner half of the artery wall. As mentioned, lipids in the blood are carried by large protein molecules, which do not diffuse into tissues. Cells in need of them have to capture them by means of surface receptors that bind lipoproteins on contact (Goldstein and Brown 1977). The receptors are located in a special area of each cell's surface, the coated pit. When receptors have captured a sufficient quantity of lipoproteins, the coated pit sinks into the cell and unites with a lysosome, the digestive organ of the cell (Simionescu, Simionescu, and Palade 1976; Anderson, Brown, and Goldstein 1977).

A brief description of the construction of the artery wall is needed at this point. The artery wall is a three-layered structure. Its inside, the intima, consists of a single layer of flat endothelial cells and a thin, liquid-filled subendothelial space. The second layer, the media, forms the bulk of the artery wall. It consists of elastic laminae and smooth muscle in varying proportions. The largest arteries, such as the aorta, are mainly elastic; those next in size, mainly muscular. The muscles of the media form peripheral bands, the contraction of which can reduce the lumen of the artery, regulating blood flow through it. Thus they reduce blood flow to organs at rest and increase it to organs in a state of high activity. However, individual smooth muscle cells also exist in the artery wall and perform the same duties as fibroblasts in other tissues. They can secrete various types of elastic and nonelastic fibers that are used mainly in the repair of injuries. Lastly, the outer layer of the wall is a tough connective-tissue covering.

The exact causes of the most important artery disease, atherosclerosis, are not known with certainty. A plausible explanation of the pathogenesis of the disorder is the following: The endothelial cells of the intima, in direct contact with blood inside the artery, capture lipoproteins in the plasma, as do other cells, by means of surface receptors. They do this not only for their own use but also for the supply of other cells in the inner half of the artery wall. Receptors in coated pits sink, as usual, into the interior of the endothelial cell, but in this case, they travel through the cell to emerge on its other side, where they discharge their load of lipoproteins into the subendothelium (Schwartz et al. 1977).

The hazard of the process, at least in prosperous countries, is not that they cannot transfer enough lipids into the subendothelium, but that they transfer too much. Presumably, if a cell captures lipoproteins from blood for its own use, it ceases to produce surface receptors when its needs are satisfied, but if it does so for other cells, it does not receive clear signals when to stop. The quantity actually transferred may depend on the availability of lipoproteins in the plasma.

In developing countries where the population lives on a barely adequate, or less than adequate, diet, the lipoprotein concentration in the plasma is correspondingly low, and the risk of endothelial cells oversupplying the cell population of the median layer of the artery wall is small. In prosperous countries, hyperlipidemia is common, presumably resulting in the almost universal occurrence of fatty suffusions in the intima of some arteries.

Such accumulations of lipoproteins, beginning in infancy or early childhood, could be the initiating event in atherosclerosis, even if the given description is only a simplified – possibly oversimplified – account of the earliest stage of atherosclerosis. Fatty streaks make a patchy appearance in the aorta and some other large arteries. The coronary circulation is usually more heavily involved than others, so that the concentration of lipoproteins in the plasma cannot be the only factor involved in pathogenesis.

Another hazard arises from the fact that arteries must be self-sealing in case of injuries; otherwise, small injuries may cause fatal hemorrhages. Injuries are sealed in the first instance by platelets – small discs 2 to 4 micrometers (μm) in diameter – circulating in immense numbers in the bloodstream. When the wall of a blood vessel is injured, collagen – a tough connective tissue inside it – is exposed. Collagen apparently attracts platelets, which immediately adhere to it. The platelets then release a substance that attracts more platelets and that makes them sticky, causing them to aggregate and form a temporary plug. This is subsequently converted into a more stable clot, after a complicated set of reactions, by fibrin. The latter is produced by the conversion of the soluble plasma protein, fibrinogen, into an insoluble substance. The fibrin originally forms a loose mesh, but ultimately it becomes a dense, tight aggregate.

This essential defense mechanism can become a life-threatening hazard, because advanced atherosclerosis can result in the ulceration of the artery wall, which, in turn, may invoke the clotting reaction. The ultimate result can be a thrombus, occluding the artery.

Atherosclerosis

Atherosclerosis begins with the already mentioned lipid suffusions – the fatty streaks – in the subendothelium of some arteries. The space immediately beneath the single layer of endothelial cells inside the artery is originally small, but it is gradually extended by lipid deposits. The accumulating lipids are mainly cholesterol esters (cholesterol with an attached fatty acid molecule). Cholesterol can act as a weak base and, combined with a weak fatty acid, constitutes the equivalent of an inorganic salt. As there are many types of fatty acids that can combine with cholesterol, there are many types of cholesterol esters. The most important esters are oleates and linoleates,

which are cholesterol combined with oleic and linoleic acid, respectively.

Cholesterol is a highly stable waxy substance, an essential constituent of every single cell in the body. The flexibility of the animal body, in comparison with the more rigid structure of plants, is due to the fact that the outer shell of plant cells is cellulose, a comparatively rigid substance, whereas that of animal cells is a complex, three-layered construction, containing proteins and lipids, including cholesterol. These cell membranes combine toughness with flexibility. In addition to the cell itself, its organelles have similar membranes.

The usefulness of cholesterol does not alter the fact that its accumulations in the artery wall can become harmful. Not known with certainty is why deposits of cholesterol esters accumulate inside artery walls. The possibility that if the density of lipoproteins in the plasma is high, endothelial cells transfer more than the required quantity from the blood to the subendothelium has been mentioned. The arising fatty streaks are visible without magnification as yellowish lines or small patches on the luminal surface of the artery. These streaks are flat, or only slightly raised, so that they do not obstruct blood flow and are not known to cause any clinical symptoms; hence, they appear to be harmless.

Fatty streaks appear in large arteries, particularly the aorta, at a very early age, probably in infancy. They are universal in children of all races in all parts of the world, but in developing countries they disappear in early childhood, whereas in prosperous countries they persist and spread. They occupy about 10 percent of the aorta in children; in adult life the affected area may increase to 30 to 50 percent.

In spite of their apparent innocuousness, fatty streaks may not be entirely harmless. They may, for instance, impede the diffusion of oxygen and water-soluble nutrients from blood flowing through the artery to tissues of the artery wall covered by them. But whatever the reason, the immune mechanism of the body treats them as foreign matter. The lipid accumulations are immediately invaded by phagocytes (monocytes), making their way into the lesions from arterial blood. They are apparently capable of separating adjacent endothelial cells slightly and slipping through the gap. Inside the fatty streak they engulf cholesterol esters until they are swollen with them. The lipid-laden phagocytes presumably attempt to leave the site of the lesion, but unlike other tissues, arteries are not drained by lymphatics, making it difficult for phagocytes to find their way out. Occasionally they can be found in the blood, and so some can escape, but many of them die in the attempt, so that ultimately dead foam cells, as lipid-laden monocytes are called, become the main constituents of fatty streaks.

Subsequently, fatty streaks are invaded by another type of cell: individual, motile smooth muscle cells,

migrating into the lesion from the media. These perform the same function as fibroblasts in other tissues. They are cells capable of secreting various elastic and nonelastic fibers, used mainly in the repair of injuries. The fibers bridge over gaps in tissues separated by the injury. As far as possible, they attempt to restore the status quo ante, but when this is not possible, they do a serviceable repair. The ends of a broken bone, for instance, would initially be held together by inelastic fibers until new bone is formed to fill the gap. Under natural conditions, the bone would heal crookedly because the repair mechanism does not have the power to reset it in its original form.

Another function of fibroblasts is to encapsulate foreign matter in the body. If the immune mechanism is powerless to deal with some invading microorganism, an attempt might be made to encapsulate its colonies, as, for instance, in tuberculosis. This is essentially the function that fibroblasts of smooth muscle origin appear to fulfill in fatty streaks. Probably they attempt to break up a continuous layer of lipids, dead phagocytes, and other cell debris into a number of separated accumulations, which are then encapsulated by layers of fibrous tissue. This is how the archetypal lesion of atherosclerosis, the fibrolipid plaque, is thought to arise. The two basic components of the fibrolipid plaque are a white cap of connective tissue and an underlying pool of necrotic debris. The proportion of the two components can vary considerably, and so can the extent of the lesion and the number of fibrous caps in it. In the advanced stage of the lesion, the fibrous caps can be so numerous that they give the luminal surface of the artery wall the porridge-like appearance from which the disorder takes its name.

The fibrolipid plaques are elevated above the surface of the intima and thus represent an obstruction to smooth blood flow. It must be remembered, however, that the lesions are virtually universal in the population of advanced countries. They begin in early childhood and last throughout life, causing no discomfort or inconvenience of any kind for decades. Indeed, in the majority of the population they remain silent until death from some other cause. They can be best conceptualized as a potentially lethal disorder, but one that is generally under the firm control of the defense mechanism of the body.

In the last, uncontrolled stage of atherosclerosis, arteries densely covered by fibrolipid plaques may ulcerate, possibly leading to the adhesion of platelets to the lesions and ultimate thrombus formation. Ulcerated arteries may calcify. The much-thickened intima of the artery wall can fissure, become partly detached, and cause the condition known as an aneurysm. If the lumen of the artery, already narrowed by fibrolipid plaques, is further obstructed by a thrombus or aneurysm, it may become completely occluded, cutting off the blood supply of tissues served by the given artery.

To repeat, the basic cause of atherosclerosis in prosperous countries is thought to be excess lipid intake, and preventive efforts throughout the century have been directed at the reduction of fats in the diet. The main difficulty has been that an effective reduction of lipids seems to require such a Spartan diet that the population of advanced countries has not been persuaded to adopt it. Much evidence, however, is available to suggest that even if this assumption is correct, it can be only a part of the story. There must be other contributory factors that determine why arterial disorders become a lethal disease in some individuals and remain in their harmless, controlled stage in others.

For example, there are considerable differences between the geographical distribution of coronary artery disease and cerebrovascular disease (strokes). The latter appears at a lower level of prosperity, the highest mortality rates occurring in moderately prosperous countries, such as Bulgaria and Portugal, where the consumption of animal fats is comparatively low. With increasing prosperity and increasing fat consumption, mortality from strokes tends to drop. Thus 40 years ago, the world leader in stroke mortality was Japan. The dramatic rise of prosperity in that country was accompanied by considerable changes in diet, the consumption of animal fats increasing by about 50 percent. At the same time, there was a sharp decrease in the incidence of strokes, and stroke mortality is now approximately half that of 40 years ago. The lipid theory of atherogenesis does not differentiate between atherosclerosis in coronary arteries and atherosclerosis in cerebrovascular arteries, and so the fall of cerebrovascular mortality in Japan (as well as a similar sharp fall in the United States in the first half of the twentieth century) provides directly contradictory evidence.

If possible, preventive measures should eliminate atherosclerosis completely, but if that is not a practical possibility, preventing the transition from the controlled to the uncontrollable stage would be nearly as useful. Therefore, an understanding of the basic process of atherogenesis is not enough. It is also necessary to investigate contributory causes and other factors that promote or retard its development. Well-known factors of this nature are cigarette smoking, diabetes, and hypothyroidism, all of which may promote atherogenesis and hasten the onset of its terminal stage, but which fall far short of providing a complete explanation. The low prevalence of coronary artery disease in Japan, for instance, has not been changed by heavy smoking habits. The effect of smoking, diabetes, and hypothyroidism is probably not specific to arterial disease, but is rather a general weakening of the immune mechanism.

Progress toward a better understanding of the causes of artery diseases can proceed along two lines. One is the study of the pathogenetic process itself. If that were completely understood, the causes might

be self-evident. Until then, important pointers may be provided by epidemiological studies. Knowledge of the parts of the earth in which artery diseases are highly prevalent and of those in which they are not should throw light on the conditions that promote or retard the disorders and to help in an understanding of their causative agents.

Epidemiological Studies

Arterial disorders have a distinct geographical distribution. They are generally very low or nonexistent in poor countries but are the leading cause of mortality in some of the most prosperous countries. Such a pattern creates the impression that the circulatory system, in spite of its complexity and potential vulnerability, is well able to cope with its natural hazards. Prosperity may have introduced some new condition – perhaps a surfeit of food in place of scarcity – that has changed a trouble-free mechanism into one susceptible to trouble.

An alternative possibility is that the difference is the result of the prolongation of life in prosperous countries. Arterial disorders appear in old age; therefore, the longer duration of life increases their share among causes of mortality. Yet another possibility is that the difference is the result of racial characteristics, Europeans being more vulnerable than Asians or Africans. Climatic conditions could be another factor, with arterial disease being more prevalent in temperate than in warm climates.

Epidemiological studies have eliminated a number of these possibilities. For instance, immigrant studies have disposed of the racial hypothesis. Black people may be free of arterial disease in Africa, but in the United States they are as vulnerable as American whites. Nor is the prolongation of life a satisfactory explanation. When mortality rates are compared in the same age groups in developing and advanced countries, it is still only the population of prosperous countries that is highly vulnerable to arterial disease. The critical factor appears to be diet. Thus, when population groups migrate from poor to prosperous countries, they, or their descendants, gradually become subject to the mortality patterns of their hosts. Migrants going from prosperous to poor countries take their diseases with them. This is, presumably, because migrants from poor to prosperous countries are willing enough to adopt the food of their hosts, but those moving in the other direction are not willing to embrace the poorer diet of their hosts.

Yet although such epidemiological studies have produced useful results on general points, they have brought little progress in matters of detail, such as the identification of food items connected with specific disorders, or finding the reason for the epidemiological peculiarities of some disorders. Why, for example, is worldwide mortality from coronary disease highest in Finland and that from strokes in Bulgaria?

One reason for the apparently limited usefulness of epidemiology is the intrinsic difficulty of the task. The diet of prosperous countries consists of something like a thousand food items, containing a million chemical substances, all potential suspects. None, however, is strongly poisonous, suggesting that perhaps the pathogenic agent may not be present in just one food item but in several, in different proportions, so that the correlation between the disease they collectively cause and one of the food items may only be weakly positive.

Possibly an equally important factor retarding progress might be regarded as a self-inflicted injury. Although the identification of the true causes of artery diseases is difficult, nothing is easier than the presentation of suspects on the basis of superficial arguments. Some of the favorite concepts of popular medicine, such as the particularly harmful effect of saturated fats or the assumed protective effect of dietary fiber, garlic, onions, and the like, are not the results of epidemiological studies but of superficial observations, or they are simply guesses.

An early comparison was that between the “low-protein, lowfat, high-fiber” diet of this or that African tribe and the “high-protein, high-fat, low-fiber” diet of prosperous countries, giving rise to the fiber industry. In fact, the differences between the diets of African tribesmen and the population of rich countries are innumerable. The tribesmen eat fewer apples, chocolates, and cabbages than do people in the West and do not take a sleeping pill at night and a tranquilizer in the morning. The main virtue of fiber is probably the fact that it is a nonfood that may help to prevent over-nutrition.

When it was pointed out that Eskimos – whose diet was not a lowfat, low protein, high fiber one – did not suffer from arterial diseases, the objection was sidestepped with the suggestion that they consumed the unsaturated fat of marine animals, not the hard fat of land animals. When it was pointed out that the French consumed as much animal fat and proteins as the British, yet the mortality from coronary disease in France was about a third of that in Britain, the “protective” effect of garlic, onions, and red wine was suggested as a possible explanation. Not even a guess is available as to how garlic, onions, and the like might protect the circulatory system. Yet such ideas, for some reason, are immediately taken up by the media and popular medicine, often impeding serious research. It may also be well to remember that the first task of research is to discover the causes of arterial disease. When those are known, attention can be turned to protective effects, but before they are known with certainty, protective effects are usually invoked to cover discrepancies between facts and theory.

A new line of approach emerged in the 1940s with the Framingham study. The basic idea was to recruit a large number of participants and keep a

record of their food consumption until they all died. Then the diet of those who died of coronary disease could be compared with that of the participants who died of other diseases, providing valuable information about the dietary causes of coronary disease. In some of the later trials, the protocol was slightly changed. Some participants were persuaded to change their habits (for example, to reduce fat intake, or to give up smoking), whereas others were allowed to continue with their customary diets and other practices. Thus, prospective studies became intervention trials. A large number of these were conducted in many countries, among others by M. R. Garcia-Palmieri and colleagues (1980), the Multiple Risk Factor Intervention Research Group (1982), and G. Rose, H. D. Turnstall-Pedoe, and R. F. Heller (1983).

Neither the Framingham trial nor its many successors, however, have produced clear evidence identifying the causes of coronary disease, and some of the results have been contradictory. At least one of the reasons for the poor performance of prospective studies is likely to be the gross inaccuracy of the food consumption data they collect. The data are obtained by means of periodic samplings by the method of “24-hour recall,” and are subject to several errors. A one-day sampling in, let us say, a month is unlikely to be truly representative of food consumption for the whole period. The data, as collected, are nonquantitative, and the quantities must be estimated by the conductors of the trial. There have to be errors of recall. These are in addition to several confounding factors, such as individual sensitivities to food toxin. If an individual dies of a food-related disorder, it is not necessarily because his consumption of the food was too high. Another possibility is that his resistance to its toxic effect was too low. Also to be remembered is that atherosclerosis is a virtually universal disorder in prosperous countries. Thus, those participants in a trial who die of other diseases may also be in an advanced stage of atherosclerosis. Finally in the case of intervention trials, the persuasion aimed at one group of participants to change habits (such as quitting smoking) is often little different from that directed at the other group by doctors and the media.

This writer suggests that food consumption statistics of advanced countries provide more reliable information regarding the type and quantity of food consumed in a community than do data collected individually in periodic samplings. Statistics are not free of errors, but they are unlikely to be of such magnitude that two studies can produce directly contradictory results. Comparing population groups rather than individuals has, at least, the advantage of eliminating individual susceptibilities. The results of a statistical study conducted by this writer (Seely 1988) is recapitulated in the following section.

A Statistical Study

When searching for the correlation between one or more food items and mortality from a given disease, statistics are likely to present a blurred image. This is partly on account of chance correlations and indirect correlations, and partly on account of statistical errors. Food consumption statistics of prosperous countries, for example, include a large amount of waste, whereas death certificates may give erroneous causes of death. Prosperity-related diseases can be positively correlated with virtually any aspect of that prosperity, such as the number of two-car families in a community, as well as with the consumption of the food item that is the presumed actual cause of disease. The consumption of a number of luxury foods increases together with prosperity. All such errors tend to mask the connection between real causes and effects. In an attempt to sharpen the image, the following expedient was tried.

Let us consider a statistical study in a group of prosperous countries, such as the 21 countries of the Organization of Economic Cooperation and Development (OECD), for which both mortality and food consumption statistics are available (Seely 1988). The object is to see if a consistent association can be found between mortality from coronary artery disease and the consumption of one or more food items. It may be advantageous to restrict the initial search to the four countries with the highest, and the four countries with the lowest, mortality from coronary disease because the contrast between the two groups may bring the critical differences between them into sharper focus. The range of mortality between the two groups is wide. Male coronary mortality in the leading country, Finland, is about eight times as high as in the country with the lowest mortality, Japan. The average male coronary mortality in the four leading countries – Finland, Ireland, United Kingdom, and Sweden – is four times as high as in the low-mortality group – Spain, France, Portugal, and Japan. Thus it seems unlikely that a food item, the consumption of which in the low-mortality group *exceeds* that in the high-mortality group, can play a significant part in the causation of coronary disease, and thus it can be excluded from the list of suspects with a reasonable degree of confidence.

In our statistical study, the source of mortality data are mortality statistics of the World Health Organization (1983–6) and that of food consumption are statistics of the Organisation for Economic Co-operation and Development (1981). Mortality statistics are for 1983, or the nearest available year; food consumption statistics are for 1973. The reason for the time interval between them is that foods containing some mildly noxious substance result in a disorder only after a long delay. The exact interval cannot be estimated, but empirical observations suggest that 10 years is reasonable.

In order to ensure that no food item for which consumption statistics are available is overlooked, it is necessary to consider *all* food appearing in them. As noted, the diet of prosperous countries includes about a thousand food items, yet the statistics give consumption figures for about 50 items. There is obviously no guarantee that foods not appearing in statistics cannot be causative agents for a disorder, but at least it can be said that all *major* food items are included in the statistics. The possible connection between a food item and a disease can be verified by the calculation of correlation coefficients. The simplest case of perfect correlation is exact proportionality between the consumption of a food item and mortality from a given disease. In other words, if in a country the consumption of x coexists with a number of deaths from a disease, y ; then in other countries, where the consumption of that food is 2, 3, or 4 times higher than in the first country, mortality is also 2, 3, or 4 times higher. In a more realistic case the relationship is $y = ax + b$, where a and b are constants, and correlation is perfect if mortality in every country corresponds exactly to the value obtained from the equation. Perfect positive correlation is expressed by the correlation coefficient 1, entire lack of correlation by the coefficient 0, and perfect negative correlation by -1 .

The results of the study are reproduced in Table IV.F.4.1. This shows the correlation coefficients calculated for each item appearing in the cited food consumption statistics for eight countries. Table IV.F.4.2 shows male age-compensated mortality rates in these countries and gives some of the more important food consumption figures from which the correlation coefficients in Table IV.F.4.1 were calculated. The highest positive correlations in Table IV.F.4.1 are as follows: oats 0.95, whole milk 0.91, milk proteins (excluding cheese) 0.91, milk fats (excluding butter and cheese) 0.91, sugar 0.90, total milk protein 0.86, beer 0.86, total milk fats 0.84, total animal fats 0.77, total animal proteins 0.74.

When correlations were checked for all 21 OECD countries, the high correlation with beer was found to be spurious. The countries with the highest beer consumption, notable Germany, Austria, the Netherlands, and Denmark, have only moderately high mortality from coronary disease. The other correlations, including that of oats, were confirmed.

The most surprising finding of the study is the near-perfect correlation between coronary mortality and the consumption of oats. At the time of the investigation, we could offer no explanation or even a suggestion as to what constituent of oats could have a connection with arterial disorders, but an attempt at an explanation follows shortly.

The connection between milk consumption and coronary disease has been suspected for a long time, based mainly on the geographical correspondence

Table IV.F.4.1. *Correlation coefficients between age-compensated male mortality rates from ischaemic heart disease and the consumption of various foods in 8 member countries of the Organization of Economic Cooperation and Development. The 8 countries comprise the 4 with the highest coronary mortality (Finland, Ireland, U.K., Sweden) and the 4 with the lowest mortality (Spain, Portugal, France, Japan). Source of mortality data: World Health Organization statistics for the year 1983 (or nearest); source of food consumption data: statistics of the Organisation of Economic Cooperation and Development for the year 1973*

Wheat	0.03	Eggs	0.11
Rye	0.33	Fish, fresh	-0.70
Barley	0.20	canned or smoked	-0.55
Oats	0.95	Molluscs, crustaceans	-0.56
Maize	0.27	Total fish	-0.77
Rice	-0.55	Fish, protein content	-0.79
Sugar	0.90	fat content	-0.77
Honey	0.44	Whole milk	0.91
Potatoes	0.30	Skimmed milk	0.61
Pulses	-0.62	Cheese	0.05
Nuts	-0.52	Dried & condensed milk	0.05
Pulses and nuts, total	-0.74	Butter	0.77
Other vegetables, protein content	-0.89	Total milk, proteins	0.86
Citrus fruit	0.26	Milk proteins without cheese	0.91
Other fresh fruit	-0.79	Total milk fats	0.84
Fruit, conserved or dried	0.62	Milk fats without butter	0.90
Total fruit	-0.55	Milk fats, without butter & cheese	0.91
Cocoa	0.22	Vegetable oils	-0.61
Beef and veal	0.45	Wine	-0.65
Pork	0.62	Beer	0.86
Mutton	0.30	Total animal proteins	0.74
Poultry	-0.36	Total animal fats	0.77
Horse meat	-0.35	Total plant proteins	-0.74
Other meats	-0.47	Total plant lipids	-0.62
Offal	0.38	Total protein	0.40
Total meat, proteins	0.33	Total fats	0.53
fats	0.55		

Source: S. Seely (1988). Reproduced from *International Journal of Cardiology*, 20 (1988) p. 185, with the permission of the publishers, Elsevier Science Publishers, Amsterdam.

between them. Figure IV.F.4.1 shows this graphically for member countries of the Organisation for Economic Co-operation and Development. Many papers have been published on the subject, notably by J. C. Annand (1961), K. A. Oster (1971), J. J. Segall (1980), and S. Seely (1981), although the authors could not agree on the mechanism of interaction between milk and the arteries. It is interesting to note from

Table IV.F.4.1 that whatever the noxious constituent of milk may be, some process in cheese manufacture appears to destroy it.

Sugar is also a long-standing suspect. However, the main objection to the suspicion is that no reasonable proposal has ever been put forward to explain why and how sugar might have an adverse effect on arteries (Yudkin and Roddy 1964).

Table IV.F.4.2. *Sample data on which Table IV.F.4.1 is based. Mortality rates below denote age-compensated male mortality rates (European standard) from ischaemic heart disease. Food consumption in grams/day*

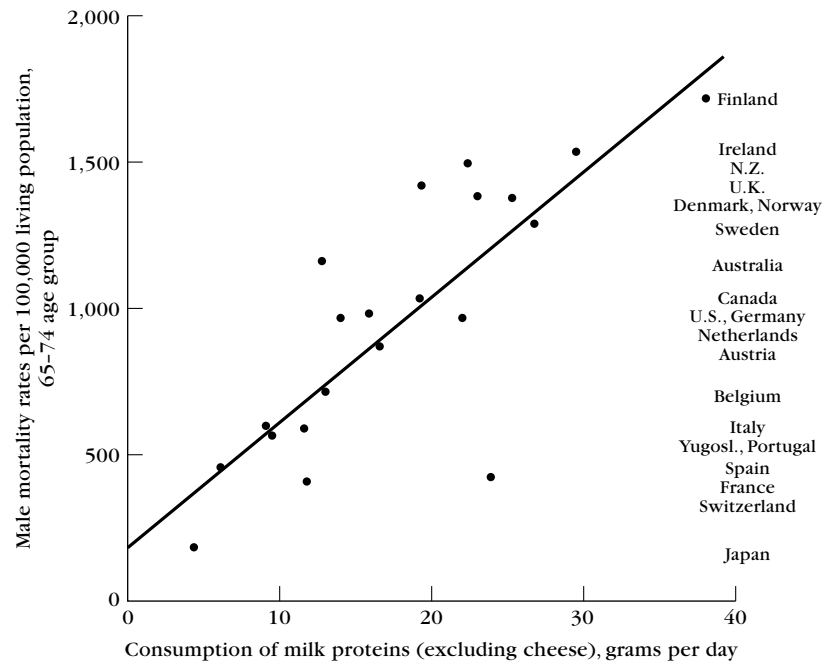
	High-mortality group				Low-mortality group			
	Finland	Ireland	U.K.	Sweden	Spain	Portugal	France	Japan
Mortality	406.4	400.1	390.6	349.2	114.6	110.3	110.1	61.4
Foods								
Wheat	125.9	221.1	177.1	151.7	207.7	197.8	192.4	84.8
Oats	7.4	6.1	3.8	5.1	-	-	0.3	-
Beef & veal	61.6	50.0	57.8	43.5	32.7	34.0	77.6	32.7
Eggs	29.4	32.1	42.1	35.0	34.6	11.9	34.5	44.1
Butter	25.5	29.6	17.1	16.0	0.9	1.4	20.1	1.2
Milk protein (without cheese)	39.4	31.8	19.3	24.9	10.2	5.5	10.9	4.0

Source: S. Seely (1988). Reproduced from *International Journal of Cardiology* with the permission of Elsevier Science Publishers.

Figure IV.F.4.1. Male coronary mortality in the 65–74 age group in OECD countries and the consumption of milk proteins (excluding cheese).

Reproduced from *International Journal of Cardiology*, with the permission of the publishers, Elsevier Science Publishers, Amsterdam.

Note. Mortality statistics are for 1987 (or the nearest year for which statistics were available); food consumption data precede mortality statistics by ten years.



The food items, for the possible atherogenic effect of which reasonable explanations *are* available, are animal fats and proteins. Yet, as shown by Table IV.F.4.1, the correlation coefficients between them and coronary mortality are considerably weaker than those found for oats, milk, and sugar. Clearly, this helps to demonstrate the complexity of the problem.

Perhaps a clue to the resolution of some of these difficulties can be found by considering one distinguishing feature that cow's milk and oats have in common: They both have a high calcium content. That of cow's milk is 120 milligrams (mg) per 100 grams (g) (as compared with 31 mg per 100 g of human milk); the calcium content of oats is 80 mg per 100 g (McCance and Widdowson 1980). Hence, we arrive at the possibility that the missing link in the pathogenesis of coronary disease is the excessively high calcium content of the Western diet.

Calcium in the Western Diet

What might be called the "natural human diet" is low in calcium, as are present-day diets in most developing countries, where the daily intake amounts to about 200 to 500 mg. The calcium content of the European diet was probably the same in past centuries, giving rise to the frequent occurrence of calcium deficiency disease – rickets – in children.

The large increase in the calcium content of the Western diet came with the increasing use of cow's milk as a staple food for all age groups. The calcium content of milk is high, and it serves the needs of infants whose rapidly growing skeletons need a com-

paratively large amount. Calves grow at 4 times the rate of human infants; hence, cow's milk contains 4 times as much calcium as human milk. The average intake of milk and dairy products in prosperous countries is usually between $\frac{1}{4}$ and $\frac{1}{2}$ liter a day, containing 300 to 600 mg calcium. Three hundred mg is sufficient to convert the intake from other foods into an abundant 600 to 700 mg, and half a liter of milk into an excessive 900 to 1,000 mg. In many Western countries, the calcium intake from cow's milk is more than that from all other foods combined.

Calcium is an essential nutrient, but in the ideal case, the dietary intake should not exceed requirements because the excretion of the surplus is a difficult task. The requirement of the human body can be estimated from the following data: The adult body contains about 1,100 grams of calcium, 99 percent of which is in the skeleton. The remaining 1 percent is needed for various essential functions, such as the generation of nerve impulses, muscular contraction, blood coagulation, and so forth. The skeleton reaches its maximum size and weight at the age of 35 years. While the skeleton is growing, it takes up a daily average of 80 mg calcium. After the age of 50 years, the skeleton begins to shrink, releasing calcium.

All body fluids contain calcium. About 100 mg is lost daily in urine and about 15 mg in sweat, though with hard physical exercise, sweating can excrete 80 mg calcium in a day. In addition, digestive fluids, such as saliva and pancreatic juice, discharge calcium into the digestive tract, and this amount is not completely reabsorbed. Allowing 150 mg calcium daily for this loss, the calcium requirement of a young adult,

depending on physical exercise, amounts to 330 to 390 mg/day, and that of an old person about 240 mg/day. The dietary intake should exceed the needs of the body by about 20 percent to allow for calcium passing unabsorbed through the intestines, making the daily requirement for a hard-working young adult 470 mg, that of old people 290 mg. When the risk of coronary heart disease is considered, we are mainly interested in the older age groups. In their case, a possible dietary calcium intake of the order of 1,000 mg is *several times* the quantity they need, and that from milk alone can be twice that amount.

The uptake of calcium from the small intestine is a controlled process, so that an excessive dietary intake does not necessarily mean that calcium finding its way into body fluids must also be excessive. Control is exercised by a vitamin D metabolite, cholecalciferol, synthesized in two steps in the liver and kidney. This metabolite is the signal for the synthesis of a carrier protein in the intestine, which is the actual transfer agent for calcium through the intestinal wall. If the calcium content of the plasma is already adequate, the synthesis of cholecalciferol is discontinued, and the excess calcium passes unabsorbed through the alimentary canal. However, newly born infants do not yet possess this control mechanism. In their case, milk sugar, lactose, facilitates the absorption of calcium from the small intestine by a simple diffusion process. Under natural conditions, this facility does not involve a health hazard because when infants are weaned, lactose disappears from their diet. In a prosperous society, however, infants are never weaned, in the sense that lactose remains in their diet from birth to death. Not only is milk, therefore, high in calcium but its lactose content also enables it to bypass the control mechanism of the body (Bronner 1987).

It might be mentioned that when milk is fermented, lactose is converted into lactic acid, a biologically inactive substance. Fermented milk and its products, such as cheese, are very high in calcium but do not provide the facility for evading the intestinal control for its absorption. This is the probable reason that the correlation between mortality from coronary disease and cheese consumption is much weaker than it is with the consumption of whole milk.

If the quantity of calcium absorbed from the intestine exceeds requirements, a good excretory mechanism exists for disposing of the excess. The kidneys normally excrete about 100 mg/day. Concentration of calcium in the urine (hypercalciuria) involves the risk of stone formation in the kidneys, but a second excretory mechanism is available to support them. The surplus calcium becomes protein bound and is excreted by the liver, not the kidneys. In individuals in prosperous countries, the concentration of protein-bound calcium in the plasma, 1.16 millimoles (mmol) per liter (l), is nearly as high as that of diffusible calcium, 1.34 mmol/l (Ganong 1987), demonstrating

that calcium intake in Western countries, indeed, tends to be excessive.

The effective excretory mechanism ensures that most of the surplus calcium is eliminated, but a small fraction escapes and is ultimately precipitated in soft tissues. Under normal conditions, this process is so slow that a large dietary calcium excess can be tolerated for decades, the calcification of soft tissues becoming a health hazard only in old age. Misuse, however, can overwhelm the excretory mechanism. In the 1950s, for example, it was customary to treat gastric ulcer patients with large quantities of milk, amounting to about 2 liters per day, until a disproportionately high mortality from coronary disease was observed among them (Briggs et al. 1960). Two liters of cow's milk contains 2.4 grams of calcium, together with lactose facilitating its absorption from the intestine. The calcium absorbed from this source by the body may amount to a daily intake of 1.8 grams, perhaps 6 times the amount needed by elderly individuals. The excretory mechanism may well be incapable of dealing with such a gross excess.

As noted, excess calcium that cannot be excreted ultimately finds its way into soft tissues. Large arteries, notably the aorta, are particularly vulnerable to calcification. The heavy calcification of the aorta in individuals who died of heart disease was already observed by the pioneers of medicine in the nineteenth century, who correspondingly called the disorder "the hardening of the arteries."

Calcium deposits in arteries have two important pathological effects. One is the calcification of atherosclerotic plaques. A recent autopsy study by A. Fleckenstein and his group (1990) has found that atherosclerotic lesions appear to attract calcium from their earliest stage onward. Fatty streaks already contain, at an average, 10 times as much calcium as the surrounding normal arterial tissue. Normal atherosclerotic plaques contain 25 times as much, advanced plaques in individuals who died of coronary disease, 80 times as much. Such advanced plaques are, in effect, calcium plaques. Calcium compounds, mainly apatite, constitute about half of their dry weight, cholesterol and its compounds about 3 percent. It is calcium that gives advanced plaques bulk and rigidity and makes them potential obstacles to blood flow.

Secondly, mural deposits of calcium in the aorta and other large elastic arteries encroach on their elasticity. As pointed out, these arteries constitute an elastic reservoir that is distended when the heart injects a volume of blood into it during systole, storing energy in its stretched elastic tissues. The contraction of the reservoir generates diastolic pressure and maintains blood flow in the circulatory system when the heart is at rest. As the heart compresses its own arteries when it contracts, its perfusion is entirely dependent on an adequate diastolic pressure.

If the elasticity of the reservoir deteriorates, an increasing systolic pressure is needed to maintain

diastolic pressure at a given value. Perfusion failure in a part of the heart occurs when the aging, partly calcified elastic reservoir cannot generate sufficient pressure to force an adequate quantity of blood through narrowed and obstructed coronary arteries. Calcification is involved both in the reduction of diastolic pressure generated by the reservoir and in the obstructions presented by advanced atherosclerotic plaques. Thus, calcium excess in Western diets may well be the most important factor in the pathogenesis of coronary artery disease (Seely 1989, 1991). If the populations of Western countries were alerted to this possibility and advised to reduce their consumption, a large reduction in mortality could well be the result.

In a recent trial (Woods et al. 1992), coronary patients were treated with magnesium sulphate with beneficial results. A possible explanation is that the excretion of the four main electrolytes – sodium, potassium, calcium, and magnesium – is an interlinked process. The most difficult task of excretion for the kidneys arises when the intake of these minerals is unbalanced, high in some, low in others. Thus in the 1960s, rats on a high-cholesterol diet also had their food unbalanced in electrolytes, with an excessive sodium-potassium and calcium-magnesium ratio (Sos 1965). The rats died of repeated, humanlike heart attacks, but their lives could be prolonged if the imbalances were moderated. Thus, if human diet has an excessive calcium content, the best remedy would be its reduction, but failing that, an increase in magnesium intake can be beneficial.

As mentioned in the section “A Statistical Study,” epidemiological studies show a positive correlation between mortality from coronary artery disease and the consumption of oats, as well as of milk. The calcium content of oats, 80 mg/100 g, is high, but the strong correlation with coronary disease would probably arise only if they also contained some substance promoting the absorption of their calcium from the intestines. An oat grass, *Trisetum flavescens*, is known to contain vitamin D₃, capable of causing calcinosis in grazing animals, but no data are available to show that this also applies to cultivated oats.

The apparent connection between mortality from coronary disease and climate has been noted. The countries with very high mortality, such as Finland, Latvia, Lithuania, and Russia, have cold climates. In warmer climates, mortality is generally lower and, in tropical countries, very low or nonexistent. This may be explained by corresponding differences in calcium excretion. In a cold climate, the amount of calcium excreted by sweating is usually small, whereas a person doing hard physical work in the tropics can lose more fluid, and possibly more calcium, in sweat than in urine.

If arterial calcification is one of the main causes of death from coronary disease – the “skeleton in the atherosclerosis closet,” as a recent article called it (Demer 1995) – this could be a blessing in disguise.

The most important source of dietary calcium is one easily identifiable food item, cow's milk – hence, arterial calcification is *preventable*. The best way of achieving such prevention would be the reduction of milk consumption, particularly by elderly people. An alternative course might be the elimination of lactose from fresh milk. As mentioned, the worldwide toll of coronary disease is about two million deaths per year. The possibility deserves careful consideration.

Stephen Seely

Bibliography

- Anderson, R. G. W., M. S. Brown, and J. L. Goldstein. 1977. Role of the coated endocytotic vesicle in the uptake of receptor-bound low-density lipoproteins in human fibroblasts. *Cell* 10: 351–64.
- Annand, J. C. 1961. Atherosclerosis: The case against protein. *Journal of the College of General Practitioners* 4: 567–96.
- Briggs, E. D., M. L. Rubenberg, R. M. O'Neal, et al. 1960. Myocardial infarction in patients treated with Sippy and other high-milk diets: An autopsy study of fifteen hospitals in the USA and Great Britain. *Circulation* 21: 532–42.
- Bronner, F. 1987. Intestinal calcium absorption: Mechanism and applications. *Journal of Nutrition* 117: 1347–52.
- Demer, L. L. 1995. A skeleton in the atherosclerosis closet. *Circulation* 92: 2029–32.
- Fleckenstein, A., M. Frey, F. Thimm, and G. Fleckenstein-Grun. 1990. Excessive mural calcium overload: A predominant causal factor in the development of stenosing coronary plaques in humans. *Cardiovascular Drugs and Therapy* 4: 1005–14.
- Ganong, W. F. 1987. *Review of medical physiology*. Thirteenth edition. Los Altos, Calif.
- Garcia-Palmieri, M. R., P. Sorlie, J. Pilotson, et al. 1980. Relationship of dietary intake to subsequent coronary heart disease incidence: The Puerto-Rico Heart Health Program. *American Journal of Clinical Nutrition* 3: 1818–27.
- Goldstein, J. L., and M. S. Brown. 1977. The low-density lipoprotein pathway and its relation to atherosclerosis. *Annual Review of Biochemistry* 46: 897–930.
- Kornitzer, M., M. Dramaix, C. Thilly, et al. 1983. Belgian heart disease prevention project: Incidence and mortality results. *Lancet* 1: 1066–70.
- Little, R. C. 1981. *Physiology of the heart and circulation*. Second edition. Chicago and London.
- McCance, R. A., and E. M. Widdowson. 1980. *The composition of foods*. London.
- Multiple Risk Factor Intervention Research Group. 1982. Multiple risk factor intervention trial: Risk factor changes and mortality results. *Journal of the American Medical Association* 248: 1465–77.
- OECD (Organisation for Economic Co-operation and Development). 1981. *Food consumption statistics, 1964–1978*. Paris.
- Oster, K. A. 1971. Plasmalogen diseases: A new concept of the etiology of the atherosclerotic process. *American Journal of Clinical Research* 2: 30–5.
- Rose, G., H. D. Turnstall-Pedoe, and R. F. Heller. 1983. U.K. heart disease prevention project: Incidence and mortality results. *Lancet* 1: 1062–6.
- Schwartz, C. J., R. J. Garrity, L. J. Lewis, et al. 1977. Arterial

- endothelial permeability to macromolecules. In *Proceedings of the fourth international symposium*, ed. G. Schettler and A. Weizel, 1–11.
- Segall, J. J. 1980. Hypothesis: Is lactose a dietary risk factor for ischaemic heart disease? *International Journal of Epidemiology* 9: 271–6.
- Seely, S. 1981. Diet and coronary disease: A survey of mortality rates and food consumption statistics of 24 countries. *Medical Hypotheses* 7: 907–18.
1988. Diet and coronary arterial disease: A statistical study. *International Journal of Cardiology* 20: 183–92.
1989. Atherosclerosis or hardening of the arteries? *International Journal of Cardiology* 22: 5–12.
1991. Is calcium excess in Western diet a major cause of arterial disease? *International Journal of Cardiology* 33: 191–8.
- Simonescu, N., M. Simonescu, and C. E. Palade. 1976. Recent studies on vascular subendothelium. *Annals of the New York Academy of Sciences* 275: 64–75.
- Sos, J. 1965. An investigation into the nutritional factors of experimental cardiopathy. In *Electrolytes and cardiovascular disease*, ed. E. Bajusz. Basel and New York.
- WHO (World Health Organization). 1983–6. *World Health Statistics Annuals*. Geneva.
- Woods, K., S. Fletcher, C. Roffe, and Y. Haider. 1992. Intravenous magnesium sulphate in suspected myocardial infarction: Results of the second Leicester Intravenous Magnesium Intervention Trial (LIMIT-2). *Lancet* 339: 344–7.
- Yudkin, J., and J. Roddy. 1964. Levels of dietary sucrose in patients with occlusive atherosclerotic disease. *Lancet* 2: 6–8.